ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
2019, Vol. 14(2) 216–237

DOI: 10.1177/1745691618805436
www.psychologicalscience.org/PPS
SAGE

Are Psychotherapies for Young People Growing Stronger? Tracking Trends Over Time for Youth Anxiety, Depression, Attention-Deficit/Hyperactivity Disorder, and Conduct Problems

John R. Weisz[1], Sofie Kuppens[2,3], Mei Yi Ng[4], Rachel A. Vaughn-Coaxum[1], Ana M. Ugueto[5], Dikla Eckshtain[6], and Katherine A. Corteselli[1]
[1]Department of Psychology, Harvard University; [2]Department of Public Health and Primary Care, KU Leuven; [3]Karel de Grote University College; [4]Department of Psychology, Florida International University; [5]Department of Psychiatry and Behavioral Sciences, McGovern Medical School, University of Texas Health Science Center at Houston; and [6]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School

Abstract
With the development of empirically supported treatments over the decades, have youth psychotherapies grown stronger? To investigate, we examined changes over time in treatment effects for four frequently treated youth mental-health problems: anxiety, depression, attention-deficit hyperactivity disorder (ADHD), and conduct disorders. We used PubMed and PsycINFO to search for randomized controlled trials (RCTs) that were published between January 1960 and May 2017 involving youths between the ages of 4 and 18 years. We also searched reviews and meta-analyses of youth psychotherapy research, followed reference trails in the reports we identified, and obtained additional studies identified by therapy researchers whom we contacted. We identified 453 RCTs (31,933 participants) spanning 53 years (1963–2016). Effect sizes for the problem-relevant outcome measures were synthesized via multilevel meta-analysis. We tracked temporal trends for each problem domain and then examined multiple study characteristics that might moderate those trends. Mean effect size increased nonsignificantly for anxiety, decreased nonsignificantly for ADHD, and decreased significantly for depression and conduct problems. Moderator analyses involving multiple study subgroups showed only a few exceptions to these surprising patterns. The findings suggest that new approaches to treatment design and intervention science may be needed, especially for depression and conduct problems. We suggest intensifying the search for mechanisms of change, making treatments more transdiagnostic and personalizable, embedding treatments within youth ecosystems, adapting treatments to the social and technological changes that alter youth dysfunction and treatment needs, and resisting old habits that can make treatments unduly skeuomorphic.

Keywords
psychotherapy, children, adolescents, youth, meta-analysis, mental health

The past half-century has seen a transformation in mental-health care, with a shift toward intervention science and the emergence of empirically supported psychological therapies. The history of that shift is marked by the publication of the earliest randomized controlled trials (RCTs) in the 1960s, the accumulation of RCTs in subsequent decades, the formation of evidence review teams and task forces in the 1990s, and the compilation and periodic updating of lists and reviews of empirically supported therapies (ESTs; see, e.g., Chambless & Ollendick, 2001; Fonagy et al., 2015; Southam-Gerow & Prinstein, 2014). The array of interventions identified as ESTs is now extensive (see Nathan & Gorman, 2015; Substance Abuse and Mental Health Services Administration, 2018; UK National Institute for

Corresponding Author:
John R. Weisz, Department of Psychology, William James Hall, Harvard University, 33 Kirkland St., Cambridge, MA 02138
E-mail: john_weisz@harvard.edu

Health and Care Excellence, 2018; Weisz & Kazdin, 2017), and efforts to boost access to these treatments are evident in diverse public and private agencies and initiatives—for example, Clark (2011); clinical guidelines and recommendations from the Agency for Healthcare Research and Quality (2018); Evidence2Success from the Annie E. Casey Foundation (2018); values-driven evidence-based practices from the California Institute for Behavioral Health Solutions (2018); evidence-based practice from the California Social Work Education Center (2012); and information from the Child Health and Development Institute of Connecticut (https://www.chdi.org/) and the National Implementation Research Network (https://nirn.fpg.unc.edu/).

These developments have been of great interest to clinical scientists and clinicians whose work focuses on childhood and adolescence (herein, *youth*), the age period in which so many mental-health problems are first identified. Indeed, nearly half of all lifetime disorders emerge by the age of 14 years (Kessler et al., 2005), and by the age of 16 at least one in three youths will have experienced a disorder, with many more suffering from serious mental-health problems that affect family life, peer connections, school functioning, and longer-term relationship and employment outcomes (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003).

Given the extent and impact of youth mental-health problems, clinical scientists and funders have invested heavily in the development and testing of youth psychotherapies, and teams of reviewers have compiled the available evidence to identify those youth therapies that qualify as ESTs. Much of this work has focused on the four forms of dysfunction that account for most youth mental-health referrals: anxiety, depression, attention-deficit/hyperactivity disorder (ADHD), and conduct problems. Recent reviews have identified ESTs in each of these four problem domains. In the most recent of these reviews, multiple ESTs at the highest level of empirical support—"well-established"—have been identified for youth anxiety (Higa-McMillan, Francis, Rith-Najarian, & Chorpita, 2016; see also related reviews on treatments for youth posttraumatic stress (Dorsey et al., 2016), youth obsessive-compulsive disorder (Freeman et al., 2014). Well-established treatments have also been identified for depression (Weersing, Jeffreys, Do, Schwartz, & Bolano, 2017), ADHD (Evans, Owens, & Bunford, 2014), and conduct problems (Kaminski & Claussen, 2017; McCart & Sheidow, 2016). Each of these reviews classifies numerous additional treatments at the second highest level of empirical support—"probably efficacious."

It is certainly good news that so many youth interventions have substantial empirical support. Note, however, that the evidence to date does not suggest that we have reached a reasonable stopping point, with no further treatment development needed in these four broad domains of youth dysfunction. This is suggested by numerous reviews noting the large number of youths who do not recover after ESTs (see, e.g., Fonagy et al., 2015; Weisz & Kazdin, 2017) and by the fact that the most comprehensive meta-analyses show RCT treatment effect sizes (ESs) for youth psychotherapies falling within the medium range (i.e., .46–.54; Weisz et al., 2017; Weisz, Weiss, Han, Granger, & Morton, 1995). Translating these values into common-language ESs (McGraw & Wong, 1992) indicates that the RCTs showed a mean probability of .63 to .65 that a youth randomly assigned to the target treatment would be better off after treatment than a youth from the control group—that is, a 13% to 15% improvement over chance (.50). Moreover, two meta-analyses of studies in which youths were randomly assigned to an EST compared with usual clinical care showed ESs in the small-to-medium range (.30 and .29, respectively; Weisz, Doss, & Hawley, 2006; Weisz, Kuppens, et al., 2013); in those comparisons, the probability that youths receiving ESTs would be better off after treatment than youths receiving usual care was .58, only modestly better than chance.

These findings suggest that, although beneficial therapies have been identified, there is room for improvement. That, in turn, suggests an important question for our field: Are our methods of developing and testing youth psychological therapies producing improvement—that is, are our methods leading to increased benefit over time, with treatment effects for commonly treated problems showing an upward trajectory in ESs across the decades? If so, such a positive trend would suggest that our approaches to building and evaluating treatments are on track for incremental benefit over time, with the prospect that treatments for commonly referred youth problems will become more and more effective with continued application of current intervention science. This would be consistent with incremental gains in other youth intervention fields—for example, in pediatric cancer treatment, in which 50 years of research has increased recovery rates from less than 30% to more than 80% (Saletta, Seng, & Lau, 2014). However, if treatment effects for some types of referred problems have not increased over the years, that might suggest a need to examine our scientific strategies in relation to those problems and perhaps to consider adjustments in some of the approaches that have been followed thus far.

Previous meta-analyses of psychotherapy research—with adults and youths—have been useful but have not provided quite the kind of fine-grained, problem-specific analysis needed for a clear picture of change over time. Most such meta-analyses have either not examined time effects or have done so in a rather global manner. As examples of the latter, Wampold et al. (1997), who focused on adult therapy, and Weisz et al. (2017), who focused on youth therapy, reported a nonsignificant overall association between study year and ES for rather broad collections of studies spanning a range of treated problems and outcome measures (including measures of outcomes not specifically targeted by the treatments being tested—e.g., session attendance, therapeutic alliance, treatment satisfaction, IQ/intellectual functioning, global functioning). Such an approach is quite useful for a broad overview. However, meta-analyses that include broad, heterogeneous collections of treated problems and outcome measures may not be as well-suited to identifying empirically and clinically important trends for different treated problems, and any trends identified might not pertain to the specific outcomes that were actually targeted in treatment. In youth psychotherapy, for example, ES trajectories might differ for the treatment of the four most commonly referred forms of dysfunction—anxiety, depression, ADHD, and conduct problems—and a focus on the specific outcomes targeted by treatment would be required for an accurate picture of the effects on the treated problems. Examining distinct trajectories for these different targeted problems on outcome measures specific to those problems would contribute precision and nuance to the picture of progress to date, informing the discussion of research priorities for the future.

With that goal in mind, we examined ES trends over time in RCTs of treatments for youth anxiety, depression, ADHD, and conduct problems. To provide a sensitive test of time trends, we extended our study search from January 1963 through May 2017, identifying 453 studies spanning more than 50 years. We examined trends in ESs across study years for each of the four target problems. Doing this within the same meta-analysis ensured consistency, with the same codes and analytic procedures applied to studies across all four problem domains. Because research practices have evolved over the years, we thought it would be important to probe the extent to which time trends in ESs might be related to, and thus qualified by, changes over time in research practices. To accomplish that objective, we considered change over time in treatment effects for the four problem domains in relation to the following characteristics of the studies.

Risk of Bias

It is possible that increasingly rigorous research has made it more difficult to produce substantial effects in RCTs—for example, by eliminating error and bias that might have inflated effects in earlier years. Alternatively, more rigorous procedures might have increased precision and sensitivity, possibly enhancing treatment ESs. We therefore examined whether time trends for each target problem might differ depending on study quality/risk of bias. We considered four risk of bias indices for which information was reported in the articles: (a) participant blinding, (b) attrition rates, (c) measure objectivity, and (d) statistical power.

Control Conditions

Studies comparing treatments to relatively weak, passive control groups (e.g., wait list) generally produce larger effects than studies using more active control groups (e.g., usual care). If the use of active control groups has increased over time, this could obscure actual improvements over time in treatment effectiveness. We investigated whether time trends for each target problem might differ depending on the control conditions used in the studies.

Clinical Diagnosis

An increased emphasis over the years on requiring study participants to meet criteria for a formal diagnosis could have led to increased sample severity over time relative to studies in which inclusion only required elevated scores on psychopathology measures. The increasing use of a diagnostic requirement over the years could have worked against an increase in treatment effects over time even if treatments were in fact improving, because more severe samples might be more difficult to treat. Alternatively, the reverse could be true, because more severe samples might have more room to improve from pre- to posttreatment than samples of youths who showed lower levels of psychopathology at baseline. Either way, it seemed important to examine time trends in ESs for each target problem in relation to whether the samples were required to meet diagnostic criteria.

Clinical Representativeness

Researchers in recent years have pressed for clinical representativeness in youth psychotherapy research, encouraging the study of clinically referred youths (rather than those recruited via ads or treated nonvoluntarily) and of treatment delivered in clinical service settings (rather than labs; e.g., Lyon & Koerner, 2016; Weisz, Krumholz, Santucci, Thomassin, & Ng, 2015). Such representativeness could reduce experimenter control, resulting in a reduced ES; alternatively, treatment effects might be enhanced with increases in the clinical relevance of the participants and settings to the treatments being provided. Whatever the direction of impact, it seemed useful to examine trends over time for each of the four target problems at differing levels of clinical representativeness.

Treatment Duration

Changes over time in the mean length of treatments might influence mean ES. Shorter treatments could be less substantive and thus less effective, or longer treatments could induce client overload or confusion and thus be less effective. With both possibilities in view, we examined time trends in ESs for the four target problems as a function of treatment duration.

Treatment Type

The question of whether the type of treatment predicts outcome has been a matter of debate in the youth and adult literature, with findings differing across reports (e.g., Miller, Wampold, & Varhely, 2008; Wampold et al., 1997; Weiss & Weisz, 1995). If the type of therapy does in fact matter, then changes in ESs over time might be influenced by temporal changes in the relative representation of various types of therapy. For example, if youth-focused behavioral treatments were especially effective (as suggested in one previous meta-analysis—see Weiss & Weisz, 1995) but behavioral approaches were to have lost popularity and declined in use over time relative to other treatment approaches, such a trend might contribute to a decline in mean ESs over time. That is, a decline in mean ESs could reflect a decline over time in the use of an especially effective treatment approach, with less effective treatments (i.e., those with a smaller ES) accounting for a larger proportion of studies in more recent years. With such possibilities in mind, we examined whether time trends in ESs for the four target problems might be conditional on the type of therapy used.

Sample Composition

Some previous findings suggest that treatment effects may differ by age or gender (e.g., Weisz et al., 1995); thus, time trends in ESs might be influenced by changes over time in the age or gender composition of study samples. Thus, we examined whether temporal changes in ESs for each of the four target problems was conditional on sample composition.

Method

Search procedure

Our search focused on peer-reviewed RCTs that tested youth psychotherapies for the domains that account for most youth mental-health referrals (see Weisz & Kazdin, 2017): depression, anxiety—this included obsessive-compulsive disorder and posttraumatic stress disorder, which were included among anxiety disorders before the publication of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, or *DSM-5* (American Psychiatric Association, 2013)—ADHD, and conduct problems. We searched PsycINFO and PubMed for RCTs published between January 1960 and May 2017. For PsycINFO, we used 21 search terms linked to psychological therapy (e.g., *psychother-, counseling*) that had been used in previous youth therapy meta-analyses, crossed with outcome-assessment topic and age-group constraints. PubMed's indexing system, MeSH, searches publishers who may use different keywords for the same concepts; we used mental disorders with the following search limits: *clinical trial, child, published in English*, and *human subjects*. We also searched reviews and meta-analyses of youth psychotherapy research, followed reference trails in the reports we identified, and obtained additional studies identified by the therapy researchers whom we contacted.

We used the following inclusion criteria: (a) participants were selected and treated for anxiety, depression, ADHD, or conduct problems (treatment targets that were vague or involving multiple problems were excluded); (b) youths were randomly assigned to treatment and control conditions, and at least one of the treatment conditions was psychotherapy (conditions involving pharmacotherapy alone or in combination with psychotherapy were excluded); (c) mean participant age was between 4 and 18 years; (d) outcome measures were specific to the treated problem (e.g., for studies targeting anxiety, only anxiety measures were included; measures of other problem areas and broader functioning were excluded from analyses to ensure that our analyses would be focused specifically on the outcomes targeted by treatment) and administered to youths in both treatment and control conditions at posttreatment; and (e) studies were published in English. Psychopathology was defined as meeting criteria for a *DSM* or International Classification of Diseases (ICD) disorder or showing elevated symptoms (e.g., clinical range scores on standardized measures of psychopathology or referred by parents or teachers for intervention); both diagnosis and elevated symptoms were included because (a) both definitions of psychopathology are common in the youth treatment outcome literature (Weisz & Kazdin, 2017); (b) youths with elevated behavioral or emotional symptoms experience serious impairment (Angold, Costello, Farmer, Burns, & Erkanli, 1999; Costello, Angold, & Keeler, 1999; Silverman & Hinshaw, 2008; Weisz, 2004); (c) such youths are commonly referred for mental-health services, with elevated symptoms prompting referral more often than formal diagnosis (Weisz, Ugueto, Cheron, & Herren, 2013); and (d) diagnostic categories and their definitions and criteria within formal systems (i.e., *DSM* and ICD) have varied markedly across the 5 decades. Figure 1 shows the search and study identification process.

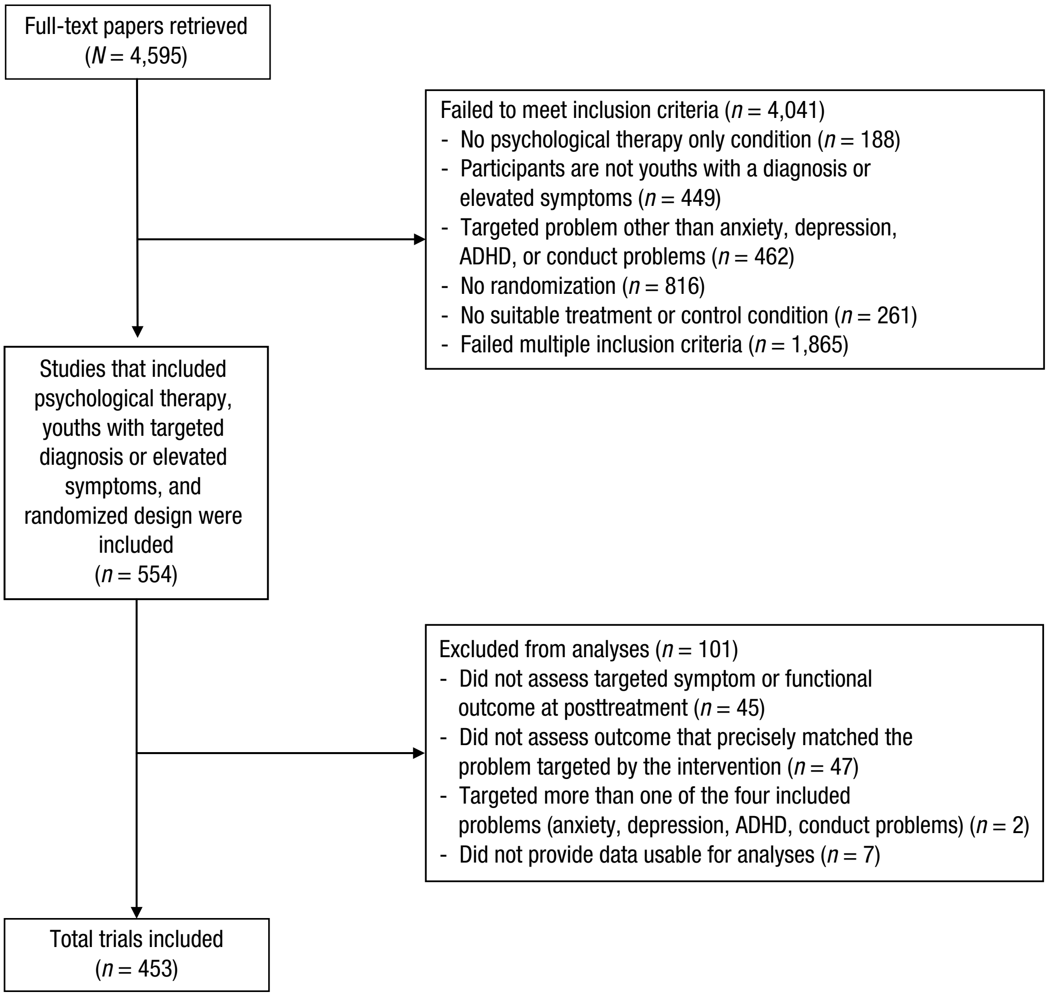

Fig. 1. Flowchart showing study retrieval, review, exclusion, and inclusion. ADHD = attention-deficit/hyperactivity disorder.

Data extraction, coding, and processing

We coded studies for multiple characteristics. The clarity of authors' reporting varied across studies, highlighting the need to assess intercoder agreement. For this purpose, seven coders each coded 20 to 30 randomly selected studies independently; the most experienced of these (an RCT researcher with a Ph.D. in clinical psychology) was the master coder against which other coders (clinical psychology postdoctoral fellows and graduate students) were compared. We included continuous codes attaining intraclass correlation coefficients (ICCs) in the "excellent" range (≥ .75) according to Cicchetti and Sparrow (1981) and categorical codes attaining kappas within Cohen's (1960) "substantial" (.61–.80) and "almost perfect" (> .80) ranges. Intercoder agreement was as follows:

1. target problem of the study—anxiety, depression, ADHD, or conduct problems (k = .97);
2. problem assessed in the outcome measure—anxiety, depression, ADHD, or conduct problems (k = .89);
3. study quality/protection against risk of bias, as indicated by participant blinding (k = .62);
4. attrition (sample size ICC = .99; then coded as "high attrition" if rate was at the 75th percentile or higher in the study pool for that target problem; range = 11%–24%);
5. objective (behavioral counts such as on-task vs. off-task behavior in class for ADHD studies or event data such as arrest counts for conduct problem studies) versus subjective (self-report or other-report, e.g., by family member or school, treatment, or research staff) measure (k = .87);
6. power (sample size ICC = .99; then coded as "adequate"—sample n = 128, providing a power of .80 to detect an ES of .50, with α = .05—vs. "inadequate");

7. control conditions (no treatment or wait list, psychotherapy or pill placebo, case management, and usual clinical care in which therapists used whatever treatments they used in their usual practice; k = .85);
8. whether a diagnosis was required for study inclusion (k = .84);
9. sample source (clinically referred vs. recruited for the study vs. receiving nonvoluntary treatment because of a court mandate or incarceration; k = .62);
10. treatment setting—clinical (outpatient, inpatient, residential, day treatment) versus nonclinical (jail, school, community, or home; k = .64);
11. treatment protocol duration in number of weeks (ICC = .95) and number of sessions (ICC = .95), both recoded into the lower-quartile (≤ Q1), interquartile (Q1–Q3), and upper-quartile (≥ Q3) sections;
12. percentage of participants who were male (ICC = .96) and White (ICC = .87), respectively, and mean age (ICC = .99), with all three then dichotomized for analysis into majority (≥ 50%) male versus female, majority (≥ 50%) White versus minority, and majority children (mean age < 12 years) versus adolescents; and
13. treatment type, collapsed into five categories: (a) youth-focused behavioral interventions (i.e., cognitive behavioral therapy, modeling, psychoeducation, operant or respondent conditioning, social skills training, biofeedback, behavioral activation, or a combination of these; the treatments could be administered individually or to groups), (b) youth-focused nonbehavioral interventions (i.e., client-centered, psychodynamic, or gestalt therapies or a combination of these), (c) caregiver and family-focused behavioral treatment (i.e., behavioral parent training, behavioral parent-youth/family interventions such as Functional Family Therapy), (d) caregiver and family-focused nonbehavioral treatment (i.e., nonbehavioral parent, parent-youth, or family interventions such as Attachment-Based Family Therapy), and (e) multisystem treatment (e.g., Multisystemic Therapy, Treatment Foster Care Oregon, or other multisystem interventions; k = .85).

In obtaining and processing the data and reporting findings, we followed Meta-Analysis Reporting Standards guidelines (American Psychological Association, 2008), with a few practical exceptions—for example, with 453 studies, it was impractical to include a table providing details of every separate study.

ES calculation

ESs were initially calculated as Cohen's d (Cohen, 1988), the standardized mean difference between treatment and control conditions on measures of the problem targeted in treatment. Our standard ES calculations used data reported in the studies or provided by the study authors whom we contacted to obtain data that were not provided in the written reports. We calculated the difference between treatment and control condition means and divided this difference by the pooled standard deviation. For studies reporting other metrics (e.g., frequencies), we transformed data to d using Lipsey and Wilson (2000) procedures. Studies reporting only p values or significant effects (assumed to reflect $p < .05$ if not otherwise stated) were assigned the minimum d that would produce that significance level given the sample size (1.81% of our cases). Studies reporting only a nonsignificant effect were assigned as $d = 0$ (Smith, 1980; 12.77% of our cases). All ES values were adjusted using Hedges's small sample correction (Hedges & Olkin, 1985), which yields an unbiased estimate of the population standardized mean difference (g).

Data synthesis

Meta-analytic approach. Because 87% of the studies yielded multiple ESs from multiple outcome measures, the assumption of ES independence was violated. We addressed the dependency via a multilevel approach that permitted the inclusion of all ESs in nonaggregated form per study. The three-level random-effects model encompassed the sampling variation for each ES (Level 1), within-study variation (Level 2), and between-study variation (Level 3). The amount of variance was expressed as the percentage of variation that lies at a particular level by estimating variance partitioning coefficients, whereas prediction intervals were computed to represent the range of the data variability. This extension of the commonly used random-effects meta-analytic model was used to examine main effects of study year on treatment ES. We also used these models to test two-way interactions between study year and other study characteristics to examine whether time trends in treatment effectiveness depended on such characteristics. In the resulting mixed-effects models, continuous or dichotomous predictors were tested using a Wald test.

For categorical predictors with more than two categories, the omnibus test follows an F distribution; pairwise comparisons were used to test which subgroup mean ESs differed significantly. Simple-slopes analyses were fitted to unravel two-way interactions by representing the time trend in treatment per subgroup of study/outcome characteristics. We used a parsimonious modeling approach, testing two-way interactions one at a time, to avoid inflating Type II error rates (Raudenbush & Bryk, 2002). Two-way interactions and simple-slopes analyses were only conducted if each subgroup contained at least five studies because parameter estimates are poor when the number of studies is very small. The percentage of explained total variance (R^2) reflected the decrease in total variance when the particular interaction was added to the model with the study year as a fixed effect. Parameters were estimated using the restricted maximum likelihood procedure implemented in SAS PROC HPMIXED, with observed ESs weighted by the inverse of the sampling variance.

Publication bias. A bias against submitting and publishing null or negative findings could inflate mean treatment effects. To investigate possible publication bias, we fitted funnel plots for each of the four problem areas using study-level mean ESs. If Egger's weighted regression test (Egger, Davey Smith, Schneider, & Minder, 1997) detected plot asymmetry, the trim-and-fill procedure (Duval & Tweedie, 2000) was used to gain insight into its potential impact. We also tested whether the odds of null or significant negative findings (i.e., with treatment group < control group) changed significantly over time for each of the four problem areas. If these publication bias indices were significantly related to the study year they could qualify the interpretation of the time-trend analyses and should thus be noted.

Results

Of the 4,595 studies retrieved and screened, 453 (for all study references, see the Supplemental Material available online) met inclusion criteria (Fig. 1); these included 31,933 participants (mean study n = 70.65; group n = 29.03). The mean age was 10.50 years (SD = 3.80), the mean percentage of males was 61.02% (SD = 25.35), and 64.57% of the study samples were White. The studies spanned from 1963 to 2016, with differences in time range per target problem (Fig. 2).

The mean posttreatment Hedges's g ES was 0.50, 95% CI = [0.45, 0.55], $t(6245) = 19.11$, $p < .001$. Between-study variance, $\sigma_v^2 = 0.242$, $\chi^2(1) = 964.1$, $p < .001$, and within-study variance, $\sigma_v^2 = 0.106$, $\chi^2(1) = 1{,}082.0$, $p < .001$, were significant, with a mean observed sampling (residual) variance of .158. Of the total variance, 47.8% was attributable to between-study differences and 21.0% to within-study differences. We found a significant two-way interaction between study year and problem targeted in the treatment on ES, $F(3, 2812) = 3.36$, $p = .018$. Simple-slopes analyses showed no significant change in ES over the years for anxiety, $b = 0.005$, $t(1093) = 1.40$, $p = .161$, or for ADHD, $b = -0.003$, $t(638) = -1.08$, $p = .281$. However, there were significant declines over time in treatment effects for depression, $b = -0.013$, $t(221) = -2.14$, $p = .034$, and for conduct problems, $b = -0.008$, $t(860) = -2.29$, $p = .022$. We focused subsequent analyses on moderators of time trends in ES effectiveness for each of the four targeted problems separately; this entailed testing two-way interactions of study year by study characteristics within each of the four target problem domains. Significant interactions were then dismantled by examining simple-slopes effects for the study year for the subgroups that were involved in the interactions.

Time trends in the treatment of anxiety

For anxiety studies, the mean posttreatment Hedges's g ES was 0.66, 95% CI = [0.56, 0.77], $t(1094) = 12.40$, $p < .001$. Between-study variance, $\sigma_v^2 = 0.334$, $\chi^2(1) = 399.1$, $p < .001$, and within-study variance, $\sigma_u^2 = 0.194$, $\chi^2(1) = 522.4$, $p < .001$, were significant, with a mean observed sampling (residual) variance of .181. Of the total variance, 47.8% was attributable to between-study differences and 27.0% to within-study differences. The prediction interval revealed that 95% of the true study mean ESs lie between −0.49 and 1.81. The overall effect of the study year on ES was nonsignificant, $b = 0.005$, $t(1093) = 1.40$, $p = .161$.

Moderators of time trends. In addition to finding no overall change in the mean ES for anxiety treatment over time, we examined two-way interactions of several variables with the study year. As Table 1 shows, only one interaction involving study year was significant (R^2 = 3.7%); it showed that the time trend in treatment effects for anxiety differed significantly according to the attrition level. Tests of simple-slopes effects indicated that measures with low attrition showed a significant increase in ESs over time, whereas there was no significant change for measures with high attrition.

The time trend in ESs for anxiety treatment did not differ significantly according to the control condition. However, we also investigated whether change in ESs might have been masked by changes in control group use by eliminating control groups altogether. We did this by examining pre- to posttreatment change within treatment groups alone. We tested whether pre-post ESs for treatment groups had changed over the years using a three-level meta-analytic model. The mean pre-post ES was large, $g = 0.84$, 95% CI = [0.66, 1.03], for the full sample of anxiety treatment studies, but with no significant change in ESs over time, $t(640) = -1.16$, $p = .248$.

Checking for publication bias. Egger's weighted regression revealed some asymmetry in the funnel plot of all study-level mean ESs for anxiety treatments (Fig. 3), $t(149) = 2.45$, $p = .016$, but the trim-and-fill procedure showed that the adjusted mean ES remained unchanged, suggesting that the impact of publication bias on the

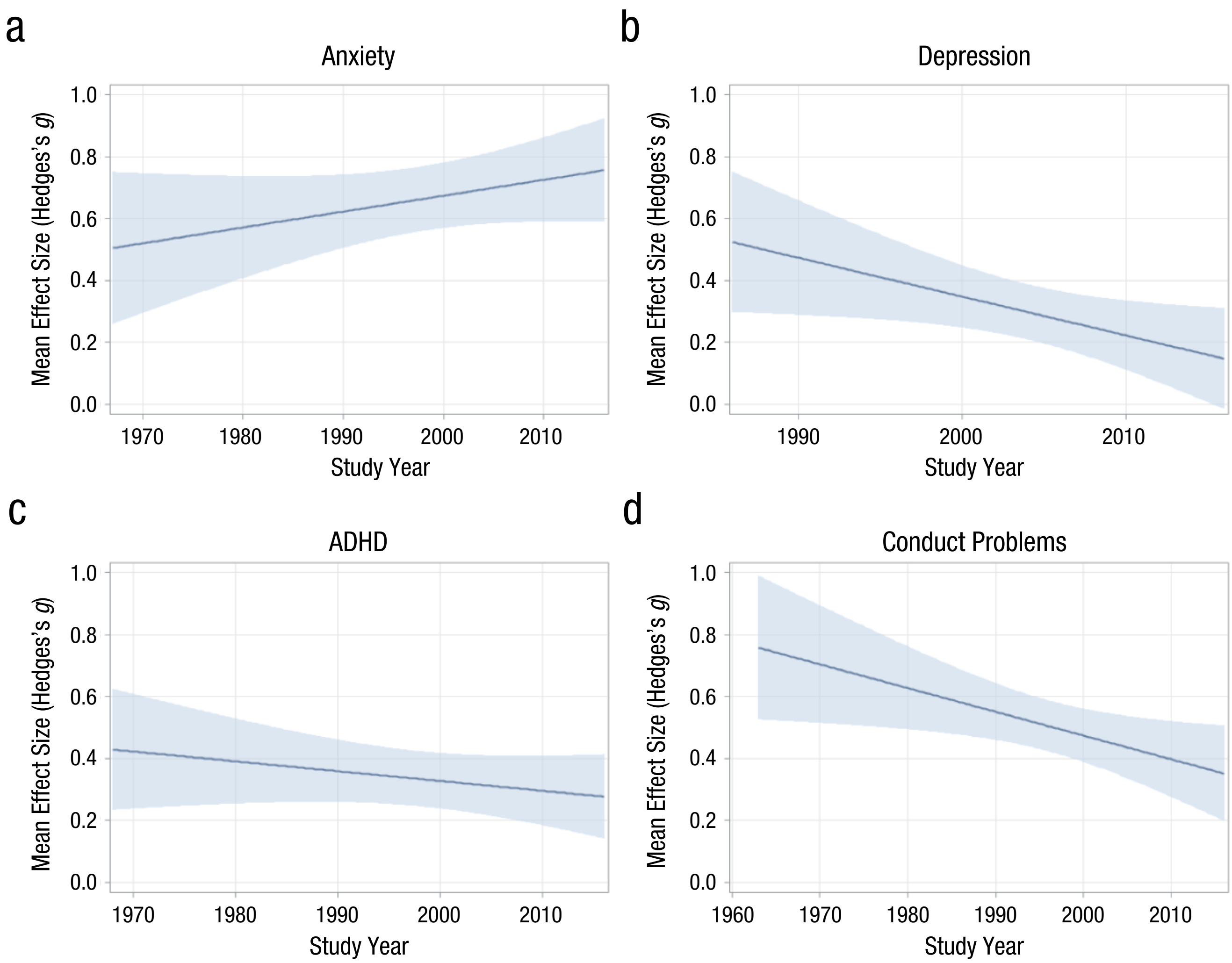

Fig. 2. Estimated change in mean effect size over time for treatment of (a) anxiety, (b) depression, (c) attention-deficit/hyperactivity disorder (ADHD), and (d) conduct problems in the mixed-effects model. The lines indicate the mean Hedges's *g*, and the blue shading around lines represents the 95% confidence interval. There was a significant interaction between study year and target problem. There was no significant change in mean effects across the years for anxiety and ADHD, but there was a significant decline for depression and conduct problems.

overall pattern of findings was minimal. In addition, there were no indications of change over time in null or significant negative findings, which could have influenced time trends in treatment ESs. That is, the odds of null findings did not change significantly over time, $t(149) = -0.79$, $p = .433$, and there were no studies of anxiety treatment reporting a significant negative finding (i.e., with treatment group < control group).

Controlling for potential confounding. We additionally examined whether adding time-varying variables to the mixed-effects models might alter the overall time trend or moderation effect. For anxiety, those variables that showed significant time trends in the multilevel logistic regression model were measure objectivity, control group, sample source, treatment setting, and treatment duration (weeks). These analyses did not substantially change the direction or magnitude of the overall time trend for anxiety and did not explain away the moderating effect.

Time trends in the treatment of depression

For depression studies, the mean posttreatment Hedges's g ES was 0.30, 95% CI = [0.20, 0.39], $t(222) = 6.27$, $p < .001$. Between-study variance, $\sigma_v^2 = 0.077$, $\chi^2(1) = 48.9$, $p < .001$, and within-study variance, $\sigma_u^2 = 0.047$, $\chi^2(1) = 45.8$. $p < .001$, were significant, with mean observed sampling (residual) variance of .101. Of the total variance, 34.3% was attributable to between-study differences, and 20.9% was attributable to within-study differences. The prediction interval revealed that 95% of the true study mean ESs lie between −0.25 and 0.84. A significant overall decline in ESs over the years was found, $b = -0.013$, $t(221) = -2.14$, $p = .034$.

Table 1. Results of Moderators of Treatment Benefit Over Time Based on Mixed-Effects Analyses of Anxiety Studies (151 Studies, 1,095 ESs)

			Simple-slopes effect for study year			Two-way interaction effect		
Subgroup	*n* Studies	*n* ESs	b_{year}[a]	*SE b*	*p*	*t* or *F*	*df*	*p*
Blinding × Study Year	151	1,095				−1.12	1091	.262
No blinding	138	841	0.006	.004	.107			
Blinding	60	254	0.000	.006	.980			
Attrition Rate × Study Year	137	968				−2.71	964	.007
Low attrition (< 17%)	113	733	0.009	.004	.022			
High attrition (≥ 17%)	41	235	−0.012	.008	.138			
Measure Objectivity × Study Year	150	1,070				−1.72	1066	.085
No objective measure	138	858	0.008	.004	.053			
Objective measure	40	212	−0.002	.006	.739			
Adequate Power × Study Year	151	1,095				1.18	1091	.240
No adequate power	147	1,078	0.004	.004	.239			
Adequate power	5	17	0.067	.049	.174			
Control Group Type × Study Year	149	1,078				1.05	2, 1072	.351
No treatment/wait list control	97	738	0.006	.004	.163			
Placebo (pill or psychological)	44	292	0.003	.006	.689			
Usual care	12	48	0.098	.068	.151			
Diagnosis × Study Year	89	718				0.27	714	.789
Not all met formal diagnosis	12	72	−0.019	.033	.568			
All met formal diagnosis	77	646	−0.009	.013	.488			
Sample Source × Study Year	132	999				−1.35	995	.177
Recruited	109	837	0.011	.005	.020			
Clinically referred	23	162	−0.011	.015	.479			
Treatment Setting × Study Year	105	728				−0.53	724	.593
Nonclinical setting	87	636	0.011	.005	.018			
Clinical setting	19	92	−0.003	.026	.898			
Planned Number of Weeks × Study Year	136	1,033				0.90	2, 1027	.408
Short (≤ 3 weeks)	36	302	−0.006	.009	.530			
Typical (4–11 weeks)	52	339	0.004	.007	.610			
Long (≥ 12 weeks)	50	392	0.012	.010	.216			
Planned Number of Sessions × Study Year	144	1,056				2.30	2, 1050	.101
Few (≤ 6 sessions)	44	352	−0.004	.007	.609			
Typical (7–15 sessions)	70	422	−0.004	.008	.634			
Many (≥ 16 sessions)	33	282	0.018	.008	.034			
Treatment Type × Study Year	151	1,095				0.30	2, 1089	.739
Youth-focused behavioral	123	796	0.005	.004	.193			
Youth-focused nonbehavioral	8	14	0.001	.015	.969			
Other	36	285	−0.003	.010	.787			
Gender Distribution × Study Year	136	1,005				−1.15	1001	.250
Majority male (≥ 50%)	52	380	0.009	.007	.168			
Minority male (< 50%)	84	625	−0.001	.005	.920			
Age Group × Study Year	151	1,095				0.29	1091	.769
Childhood (< 12 years)	101	775	0.004	.005	.335			
Adolescence (≥ 12 years)	50	320	0.007	.006	.297			

Note: Some variables were missing for certain studies. Each study can contribute multiple ESs; thus, study sample size across subgroups can exceed the total study sample size for the outcomes characteristics. ES = effect size.
[a]Simple-slopes effects for study year on treatment effectiveness for relevant subgroups. The coefficient corresponds to the change in ES with a 1-year increase; positive values indicate an increase in ES over time, and negative values indicate a decline over time.

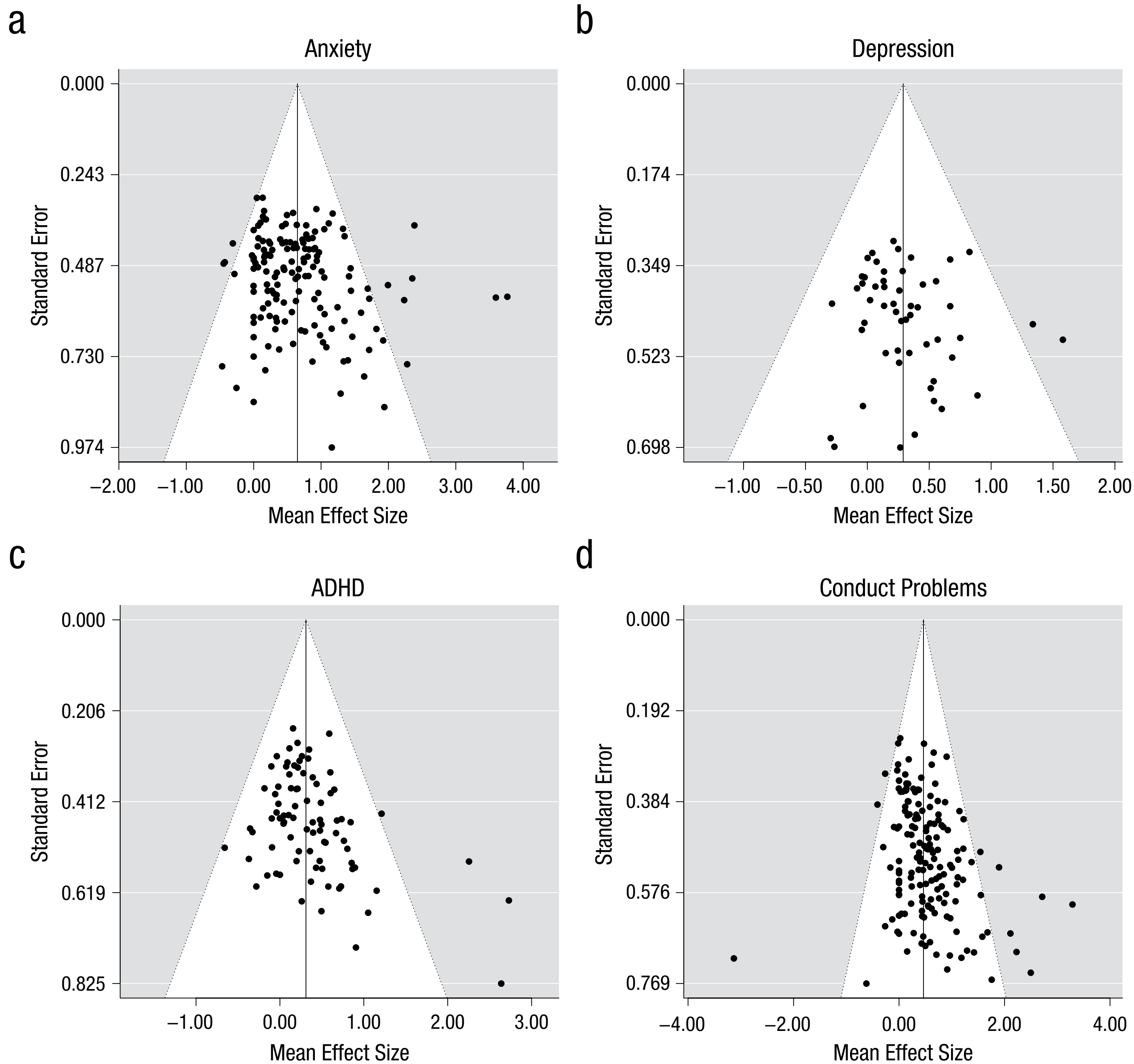

Fig. 3. Funnel plots of study-level mean effect sizes for (a) anxiety, (b) depression, (c) attention-deficit/hyperactivity disorder (ADHD), and (d) conduct problems. In the absence of publication bias, effect sizes are spread symmetrically on both sides of the overall mean effect size (vertical line) in the white area, and effect sizes in the gray area indicate asymmetry.

Moderators of time trends. To delve more deeply into the significant overall decline in mean ES for depression treatment over time, we tested for two-way interactions of time with other variables. As shown in Table 2, two interactions that involved the study year were significant, R^2 = 23.9% and R^2 = 20.2%, respectively; both interactions involved the clinical representativeness of the studies. Simple-slopes effects indicated that effects grew larger over the years for studies in which youths had been clinically referred and studies in which treatment was carried out in clinical settings, although only the former was significant. By contrast, effects declined significantly over time for studies not conducted with clinically referred samples and those not conducted in clinical settings.

There was no significant interaction of study year × control condition (Table 2), but we tried the additional approach of eliminating control groups altogether and using our three-level meta-analytic model to test for pre- to posttreatment change within treatment groups alone. The mean pre-post *g* was 0.60, 95% CI = [0.41,

Table 2. Results of Moderators of Treatment Benefit Over Time Based on Mixed-Effects Analyses of Depression Studies (53 Studies, 223 ESs)

Subgroup	*n* Studies	*n* ESs	Simple-slopes effect for study year			Two-way interaction effect		
			b_{year}[a]	*SE b*	*p*	*t* or *F*	*df*	*p*
Blinding × Study Year	53	223				1.47	219	.143
No blinding	51	192	−0.016	.006	.011			
Blinding	16	31	−0.001	.011	.929			
Attrition Rate × Study Year	49	211				1.83	207	.068
Low attrition (< 24%)	45	156	−0.018	.007	.009			
High attrition (≥ 24%)	10	55	0.004	.012	.737			
Measure Objectivity × Study Year								
No objective measure	21	218	−0.013	.006	.032			
Objective measure[b]								
Adequate Power × Study Year	53	223				0.65	219	.518
No adequate power	39	172	−0.009	.006	.119			
Adequate power	15	51	0.001	.015	.929			
Control Group Type × Study Year	53	19				0.95	2, 213	.388
No treatment/wait list control	25	103	−0.010	.008	.216			
Placebo (pill or psychological)	11	46	−0.014	.013	.291			
Usual care	19	70	0.010	.020	.445			
Diagnosis × Study Year	33	142				0.41	138	.683
Not all met formal diagnosis	15	49	−0.007	.010	.480			
All met formal diagnosis	18	93	0.000	.011	.977			
Sample Source × Study Year	51	214				3.82	210	< .001
Recruited	32	150	−0.026	.006	< .001			
Clinically referred	19	64	0.022	.011	.044			
Treatment Setting × Study Year	39	142				3.08	138	.003
Nonclinical setting	21	76	−0.022	.008	.007			
Clinical setting	18	66	0.024	.013	.058			
Planned Number of Weeks × Study Year	45	191				2.28	2,185	.105
Short (≤ 6 weeks)	13	50	−0.006	.009	.530			
Typical (7–11 weeks)	12	66	0.004	.007	.610			
Long (≥ 12 weeks)	20	75	0.013	.010	.216			
Planned Number of Sessions × Study Year	44	185				0.60	2, 179	.548
Few (≤ 10 sessions)	17	48	−0.036	.015	.018			
Typical (11–19 sessions)	23	90	−0.021	.012	.087			
Many (≥ 20 sessions)	8	47	−0.009	.021	.665			
Treatment Type × Study Year	53	223				−0.10	219	.921
Youth-focused behavioral	32	134	−0.012	.007	.065			
Other	28	89	−0.013	.008	.112			
Gender Distribution × Study Year	53	223				0.37	219	.715
Majority male (≥ 50%)	11	35	−0.016	.012	.173			
Minority male (< 50%)	42	188	−0.011	.007	.108			
Age Group × Study Year	53	223				0.74	219	.460
Childhood (< 12 years)	9	27	−0.023	.015	.120			
Adolescence (≥ 12 years)	44	196	−0.011	.007	.091			

Note: Some variables were missing for certain studies. Each study can contribute multiple ESs; thus, study sample size across subgroups can exceed the total study sample size for the outcomes characteristics. ES = effect size.
[a]Simple-slopes effects for study year on treatment effectiveness for relevant subgroups. The coefficient corresponds to the change in ES with a 1-year increase; positive values indicate an increase in ES over time, and negative values indicate a decline over time. [b]This analysis could not be performed because there were insufficient studies to yield a reliable estimate.

0.78], for the full sample of depression studies, and there was significant temporal change, $t(169) = -3.13$, $p = .002$, but with ESs *declining* over time.

Checking for publication bias. Egger's weighted regression revealed no significant asymmetry in the funnel plot of all study-level mean ESs for depression treatments (Fig. 3), $t(51) = 1.04$, $p = .305$, and the trim-and-fill procedure showed that the adjusted mean ES remained unchanged. In addition, the odds of null findings did not change significantly over time, $t(51) = 1.12$, $p = .270$, and no depression studies reported a significant negative finding (i.e., with treatment group < control group). These findings do not indicate an impact of publication bias.

Controlling for potential confounding. We examined whether adding time-varying variables to the mixed-effects models would alter the overall time trend or moderation effect. For depression, only the control group variable showed a significant time trend. The additional analyses with the control group added did not explain away the moderating effects, but the negative overall time trend became less pronounced and nonsignificant, $b = -0.007$, $t(218) = -1.12$, $p = .263$.

Time trends in the treatment of ADHD

For ADHD studies, the mean posttreatment Hedges's g ES was 0.33, 95% CI = [0.24, 0.42], $t(639) = 7.38$, $p < .001$. Between-study variance, $\sigma_v^2 = 0.122$, $\chi^2(1) = 82.6$, $p < .001$, and within-study variance, $\sigma_u^2 = 0.069$, $\chi^2(1) = 145.1$, $p < .001$, were significant, with a mean observed sampling (residual) variance of .177. Of the total variance, 33.2% was attributable to between-study differences and 18.6% to within-study differences. The prediction interval revealed that 95% of the true study mean ESs lie between −0.35 and 1.02. The overall effect of study year on ESs was nonsignificant, $b = -0.003$, $t(638) = -1.08$, $p = .281$.

Moderators of time trends. As noted above, we found no overall temporal change in the mean ES for ADHD treatment. To determine whether the change in treatment benefit might be found under certain conditions, we tested the interactions shown in Table 3. Of these, the only significant interaction was Measure Objectivity × Study Year ($R^2 = 1.0\%$). Simple-slopes effect tests showed no evidence of temporal change for outcomes assessed by subjective (e.g., self-report) measures, but more objective measures showed a significant decline in ESs over time.

To complement our control conditions analysis, we removed control groups entirely using our three-level meta-analytic model to test whether treatment groups showed temporal change in pre-post ESs. The mean pre-post ES was in the medium range, $g = 0.42$, 95% CI = [0.28, 0.56], for the full sample of ADHD studies, with no significant change in ESs over time, $t(456) = -1.18$, $p = .239$.

Checking for publication bias. Egger's weighted regression revealed significant asymmetry in the funnel plot of all study-level mean ESs for ADHD treatments, $t(83) = 3.10$, $p = .003$ (see Fig. 3). The trim-and-fill procedure estimated that 12 studies were missing on the left side of the plot. Adding these studies adjusted the mean Hedges's g downward by about 8 points, suggesting that publication bias could have affected the findings. Nonetheless, neither the odds of null findings nor the odds of significant negative findings changed significantly over time, $t(83) = 0.53$, $p = .597$ and $t(83) = 0.64$, $p = .525$, respectively; thus, there was no indication of change over time in the rates of null or significant negative findings that could have influenced time trends in ESs.

Controlling for potential confounding. We also examined whether adding time-varying variables to the mixed-effects models might alter the overall time trend or moderation effect. For ADHD, those variables were blinding, measure objectivity, control group, and treatment duration (weeks, sessions). These additional analyses did not substantially change the direction or magnitude of the overall time trend or explain away the moderating effect.

Time trends in the treatment of conduct problems

For conduct-problem studies, the mean posttreatment Hedges's g ES was 0.50, 95% CI = [0.42, 0.59], $t(861) = 11.98$, $p < .001$. Between-study variance, $\sigma_v^2 = 0.217$, $\chi^2(1) = 282.9$, $p < .001$, and within-study variance, $\sigma_u^2 = 0.083$, $\chi^2(1) = 438.5$, $p < .001$, were significant, with a mean observed sampling (residual) variance of .127. Of the total variance, 50.8% was attributable to between-study differences, and 19.4% was attributable to within-study differences. The prediction interval revealed that 95% of the true study mean ESs lie between −0.41 and 1.42. A significant decline in ESs over the years was obtained, $b = -0.008$, $t(860) = -2.29$, $p = .022$.

Moderators of time trends. The significant overall decline in mean ES for conduct problems was qualified by two significant interactions of time with other variables (see Table 4). First, an attrition rate × study year interaction ($R^2 = 1.9\%$), when broken down into simple slopes, indicated that there was a significant decline in ESs for measures low in attrition (less than 17%, bottom quartile) but no significant

Table 3. Results of Moderators of Treatment Benefit Over Time Based on Mixed-Effects Analyses of ADHD Studies (85 Studies, 640 ESs)

			Simple-slopes effect for study year			Two-way interaction effect		
Subgroup	*n* Studies	*n* ESs	b_{year}[a]	*SE b*	*p*	*t* or *F*	*df*	*p*
Blinding × Study Year	85	640				0.65	636	.513
No blinding	54	378	−0.004	.003	.190			
Blinding	56	262	−0.002	.004	.624			
Attrition Rate × Study Year	66	468				1.13	464	.258
Low attrition (< 11%)	56	335	−0.009	.004	.027			
High attrition (≥ 11%)	13	133	0.002	.009	.834			
Measure Objectivity × Study Year	85	638				−2.49	634	.013
No objective measure	57	316	0.002	.004	.631			
Objective measure	49	322	−0.008	.003	.022			
Adequate Power × Study Year	85	640				−0.02	636	.981
No adequate power	80	583	−0.002	.003	.433			
Adequate power	5	57	−0.003	.022	.894			
Control Group Type × Study Year	76	577				2.51	2, 621	.082
No treatment/wait list control	41	257	0.001	.004	.703			
Placebo (pill or psychological)	42	320	−0.010	.004	.016			
Usual care	7	50	−0.003	.020	.890			
Diagnosis × Study Year								
Not all met formal diagnosis[b]								
All met formal diagnosis	42	343	0.007	.006	.279			
Sample Source × Study Year	74	548				−1.23	544	.220
Recruited	57	426	−0.002	.004	.650			
Clinically referred	17	122	−0.011	.007	.112			
Treatment Setting × Study Year	54	429				1.20	425	.231
Nonclinical setting	42	346	−0.008	.004	.036			
Clinical setting	41	83	0.004	.010	.691			
Planned Number of Weeks × Study Year	74	507				0.16	2, 501	.853
Short (≤ 4 weeks)	21	136	0.000	.007	.998			
Typical (5–11 weeks)	33	240	−0.000	.006	.977			
Long (≥ 12 weeks)	20	131	0.008	.013	.558			
Planned Number of Sessions × Study Year	79	583				0.30	2, 577	.743
Few (≤ 6 sessions)	23	176	0.000	.006	.970			
Typical (7–30 sessions)	44	261	−0.000	.005	.973			
Many (≥ 31 sessions)	12	146	0.015	.020	.434			
Treatment Type × Study Year	85	640				2.24	2, 634	.108
Youth-focused behavioral	37	321	−0.003	.004	.388			
Caregiver/family-focused behavioral	11	63	0.026	.015	.088			
Other	44	256	−0.007	.004	.105			
Gender Distribution × Study Year								
Majority male (≥ 50%)	73	556	−0.002	.004	.666			
Minority male (< 50%)[b]								
Age Group × Study Year	85	640				−0.32	636	.746
Childhood (< 12 years)	76	524	−0.003	.003	.414			
Adolescence (≥ 12 years)	9	116	−0.006	.011	.573			

Note: Some variables were missing for certain studies. Each study can contribute multiple ESs; thus, study sample size across subgroups can exceed the total study sample size for the outcome characteristics. ADHD = attention-deficit/hyperactivity disorder; ES = effect size.
[a]Simple-slopes effects for study year on treatment effectiveness for relevant subgroups. The coefficient corresponds to the change in ES with a 1-year increase; positive values indicate an increase in ES over time, and negative values indicate a decline over time. [b]This analysis could not be performed because there were insufficient studies to yield a reliable estimate.

Table 4. Results of Moderators of Treatment Benefit Over Time Based on Mixed-Effects Analyses of Conduct Problem Studies (164 Studies, 862 ESs)

Subgroup	*n* Studies	*n* ESs	Simple-slopes effect for study year b_{year}[a]	*SE b*	*p*	Two-way interaction effect *t* or *F*	*df*	*p*
Blinding × Study Year	164	862				0.50	858	.619
No blinding	73	279	−0.009	.004	.046			
Blinding	134	583	−0.007	.004	.065			
Attrition Rate × Study Year	141	718				1.98	714	.049
Low attrition (< 17%)	115	529	−0.012	.004	.002			
High attrition (≥ 17%)	39	189	0.001	.006	.867			
Measure Objectivity × Study Year	149	802				−2.67	798	.008
No objective measure	141	693	−0.007	.004	.063			
Objective measure	31	109	−0.025	.007	< .001			
Adequate Power × Study Year	164	862				0.85	858	.394
No adequate power	144	776	−0.006	.004	.073			
Adequate power	20	86	0.008	.016	.638			
Control Group Type × Study Year	164	862				0.87	3, 854	.454
No treatment/wait list control	89	504	−0.003	.005	.487			
Placebo (pill or psychological)	25	91	−0.019	.010	.056			
Case management	27	147	−0.010	.010	.144			
Usual care	25	120	−0.003	.014	.821			
Diagnosis × Study Year	36	268				−0.32	264	.749
Not all met formal diagnosis	16	99	0.002	.0127	.850			
All met formal diagnosis	20	169	−0.004	.0158	.796			
Sample Source × Study Year	155	816				0.08	2, 810	.927
Recruited	84	398	−0.006	.005	.155			
Nonvoluntary	32	194	−0.009	.007	.195			
Clinically referred	39	224	−0.005	.007	.513			
Treatment Setting × Study Year	119	607				0.41	603	.682
Nonclinical setting	86	441	−0.015	.005	.004			
Clinical setting	34	166	−0.010	.011	.373			
Planned Number of Weeks × Study Year	134	689				0.27	2, 683	.763
Short (≤ 7 weeks)	37	183	−0.013	.009	.141			
Typical (8–18 weeks)	73	330	−0.006	.007	.385			
Long (≥ 19 weeks)	25	176	−0.012	.009	.161			
Planned Number of Sessions × Study Year	124	658				1.77	2, 652	.171
Few (≤ 9 sessions)	41	182	−0.002	.008	.778			
Typical (10–21 sessions)	62	306	−0.002	.008	.723			
Many (≥ 22 sessions)	25	170	−0.020	.008	.013			
Treatment Type × Study Year	164	862				0.96	4, 852	.429
Youth-focused behavioral	53	240	−0.015	.006	.012			
Youth-focused nonbehavioral	14	50	0.005	.010	.621			
Caregiver/family-focused behavioral	60	290	−0.005	.006	.360			
Multisystemic	12	72	−0.012	.020	.564			
Other	44	210	−0.010	.006	.062			
Gender Distribution × Study Year	151	819				−1.05	815	.295
Majority male (≥ 50%)	137	770	−0.004	.004	.245			
Minority male (< 50%)	14	49	−0.014	.009	.112			
Age Group × Study Year	162	855				−0.34	851	.736
Childhood (< 12 years)	106	555	−0.008	.005	.061			
Adolescence (≥ 12 years)	56	300	−0.011	.005	.042			

Note: Some variables were missing for certain studies. Each study can contribute multiple ESs; thus, study sample size across subgroups can exceed the total study sample size for the outcome characteristics. ES = effect size.

[a]Simple-slopes effects for study year on treatment effectiveness for relevant subgroups. The coefficient corresponds to the change in ES with a 1-year increase; positive values indicate an increase in ES over time, and negative values indicate a decline over time.

trend for higher attrition measures. Second, a Measure Objectivity × Study Year interaction (R^2 = 3.0%) reflected the fact that there was a significant temporal decline in ESs for objective measures and no significant trend for subjective measures.

There was no significant interaction of study year by control condition, but to provide a further test of whether temporal change in ESs might have been masked by changes in control group use, we eliminated control groups altogether and focused on pre- to post-treatment change within treatment groups alone using our three-level meta-analytic model. The mean pre-post ES was in the medium-to-large range, g = 0.65, 95% CI = [0.51, 0.78], for the full sample of conduct problem studies, but there was no significant temporal change, t(527) = −0.05, p = .961.

Checking for publication bias. Egger's weighted regression revealed significant plot asymmetry, t(162) = 4.26, p < .001 (see Fig. 3). The trim-and-fill procedure estimated that 42 studies were missing on the left side of the plot. Adding these missing studies adjusted the mean Hedges's g downward by about 18 points, suggesting that publication bias could have affected the findings. However, neither the odds of null findings nor the odds of significant negative findings changed significantly over time, t(162) = 1.13, p = .260, and t(162) = 1.37, p = .173, respectively, so there was no indication of temporal changes in the rates of null or significant negative findings that could have influenced time trends in ESs.

Controlling for potential confounding. As was done for the other three targeted problems, we examined whether adding time-varying variables to the mixed-effects models would alter the overall time trend or moderation effects. For conduct problems, those variables included power, control group, treatment duration (weeks), and sample age group. These additional analyses did not substantially change the direction or magnitude of the overall time trend or explain away the moderating effects.

Discussion

One way to evaluate progress for various forms of intervention science is to examine change over time in outcomes. In some fields—pediatric cancer treatment, for example—that approach has revealed striking gains over time (Saletta et al., 2014). We examined change over time in youth psychotherapy outcomes, focusing on RCTs of treatments for four frequently treated forms of youth psychological dysfunction—anxiety, depression, ADHD, and conduct problems. We found that the time trend in treatment ESs differed across the four problem domains, with no significant time trend for anxiety or ADHD and significant declines over time in ESs for depression and conduct problems. Because any such findings could reflect changes in research design and study procedures across the decades, we carried out subsequent moderator analyses to determine whether trends within each of the four problem domains might have been conditional on variations in study characteristics.

Moderators of time trends for separate target problems

For anxiety, we found no significant change in ESs over time. Moderator tests showed a significant interaction with measure attrition: ESs increased significantly over the years for measures low in attrition, with no significant change for measures higher in attrition. Our analyses of the depression studies showed a significant decline in ESs over time. That decline was qualified by two significant interactions involving the clinical representativeness of studies. Simple-slopes tests for depression showed that ESs increased significantly over the years for studies with clinically referred youths and marginally (p = .058) for studies in clinical settings but decreased over the years in studies with nonreferred youths and in nonclinical settings. It is possible that the positive trends resulted in part from what researchers learned by conducting a relatively large proportion of their trials (compared with trials for other problem domains) with referred youths (37%) and in clinical settings (46%). Depressive disorders have relatively low prevalence in young people (see, e.g., Kessler, Avenevoli, Costello, et al., 2012; Merikangas et al., 2010), and that may have led depression researchers to seek study participants by turning to clinical settings in which youths with significant levels of depressive symptoms are more likely to be found than in the general population. Whatever the reasons, the findings do suggest, encouragingly, that the effects of youth depression treatment may have grown stronger over time in tests that have involved clinically representative youths and settings.

Turning to externalizing problems, our analyses of ADHD treatment studies showed no significant change in ESs across the years, and our moderator tests revealed only one significant interaction: Measure Objectivity × Study Year. Simple-slopes tests showed no evidence of temporal change when subjective measures were used, but with objective measures there was a significant decline in ESs over time. Conduct problem treatment showed a significant decline in ESs over time. Moderator tests showed two interactions: (a) Measurement Attrition × Study Year, with a significant decline in ESs for measures showing low attrition (< 17%) but no significant change for other measures; and (b) Measure

Objectivity × Study Year, with a significant decline in ESs over time for objective measures and no significant change for subjective measures. The uniformity of the decline for conduct problem treatment was surprising given the long history of positive RCT findings (see, e.g., Kaminski & Claussen, 2017; McCart & Sheidow, 2016), but of course change in treatment effects over time has not received much attention in previous reviews.

Strengths and limitations

Certain strengths of this meta-analysis should be noted. To our knowledge, it includes the largest study pool ever assembled for an analysis of outcomes specific to the four target problem domains, and it spans the longest time period to date—thus providing a particularly robust test of change over time in treatment effects. In addition, examining trends for the four primary problem areas within the same meta-analysis rather than comparing across separate analyses made it possible to apply exactly the same coding and analytic methods across problem domains, creating a level playing field for comparison across areas. We used a state-of-the art method to address the ES dependency that is present in virtually every meta-analysis (violating the assumption of ES independence) by applying a multilevel approach that permits the inclusion of all ESs in nonaggregated form for each study (rather than, e.g., including only one ES per study or averaging ESs within studies) and modeling study and outcome characteristics in moderator analyses. To enhance the comprehensiveness of our analyses, our search included not only the primary outcome paper for each study but also other articles reporting relevant data. More broadly, our analyses provide a more detailed and nuanced picture than more global previous reports on strength of association between study effects and study year for heterogeneous samples (e.g., Wampold et al., 1997; Weisz et al., 2017) because we focused on temporal trends for separate targeted problems and then examined the extent to which those trends might be conditional on study characteristics, which might themselves change over time.

Despite these strengths, several limitations warrant attention, some resulting from characteristics of the study pool. For example, study descriptions varied in clarity, affecting intercoder kappas for some categorical codes; although all the kappas met Cohen's (1960) standard for substantial or almost perfect agreement, three codes (participant blinding, sample source, and treatment setting) were relatively low, suggesting a need for caution in interpreting related findings. In addition, the number of studies in some subgroups was too limited to yield fully reliable subgroup effects, and in other cases the n per subgroup was adequate, but unbalanced designs may have reduced power; thus, it is possible that clinically important subgroup differences in time trends went undetected. Although more pronounced for depression studies, moderators of time trends generally explained a small proportion of the total variance in ESs, suggesting that other moderators may account for the remaining variance in ESs. Because the year of study was not randomly assigned, time-related trends in ESs are vulnerable to threats to internal validity. Although we examined the impact of several time-varying variables, we cannot rule out the possibility that we may have failed to include other potential confounders that could explain the obtained time trends or could be suppressing trends showing improvement in treatment effects over time.

Other potential limitations relate to our inclusion/exclusion criteria. Inclusion of only English-language studies may have limited the generalizability of the findings. In addition, our inclusion of only published studies could be viewed as a limitation. However, the fact that our pool of studies spanned 5 decades to provide a sensitive test of temporal trends created a study inclusion challenge: Because it would have been much easier to identify unpublished work from more recent decades than from earlier decades, including unpublished studies could also have introduced bias into the study pool.

We investigated whether bias favoring publication of positive findings might have influenced overall ES results for any of the four problem areas. Egger's weighted regression (Egger et al., 1997) and the trim-and-fill procedure (Duval & Tweedie, 2000) suggested a minimal impact of publication bias on analyses of anxiety and depression but a potential impact on ADHD and conduct problem findings. However, we did not find evidence of change over time in the publication of null or significant negative findings that could have influenced our findings on temporal trends in ESs for anxiety, depression, ADHD, or conduct problems.

Implications and future research directions

Taken together, our analyses across the four problem domains revealed only three subgroup trends showing significantly improved ESs over time, and these were confined to internalizing problems. Even if we examine all the simple-slopes effects reported in Tables 1 through 4, regardless of whether they were components of significant interactions, we find only three additional anxiety treatment subgroups showing increasing ESs, no additional subgroups for depression, and no subgroups at all showing increased ESs over time for ADHD or conduct problems. Moreover, we found significant declines in ESs, year over year, for depression and conduct problem treatment, patterns echoed in 71% of the subgroup simple-slopes trends for depression in Table 2 and 88% of the simple slopes for conduct problems in Table 4. In sum, there were strikingly few exceptions to the general pattern that treatment effects were either unchanged or declining across the decades for each of the target problems. One possible implication is that the research strategy used over the past 5 decades, the treatment approaches investigated, or both, may not be ideal for generating incremental benefit over time. Several ideas may warrant attention as we seek to understand the general pattern of findings and ponder future directions.

Mechanisms. One possible explanation, suggesting one future direction, is that boosting psychotherapy effects may require that we understand what mechanisms of change are required for genuine improvement. The requirements for establishing true mechanisms are quite daunting, and most experts agree that we have little evidence to date of the kind needed to accomplish that objective (see, e.g., Kazdin, 2007). Our limited understanding of mechanisms may encourage default repetition of standardized treatments that have "worked" previously, and this may constrain innovation; repeating relatively similar therapies year after year may impose a natural limit on how much therapy benefit can increase over time. Several methods have been proposed for identifying mechanisms (Kazdin, 2007). One example, the "experimental therapeutics" approach (National Institute of Mental Health, 2016), involves testing specific intervention procedures for their ability to affect proposed mechanisms, or "targets," and then testing their ability to affect treatment outcomes via a change in the targets. Identifying true change mechanisms—through this or other methods—might help smooth the path to incremental improvements by telling us what switches need to be flipped for genuine increments in treatment benefit.

Treatment structure. Another possible explanation for our failure to find much improvement in benefit may lie in the structure of youth psychotherapies. These are typically standardized protocols containing 10 to 20 preplanned sessions that are delivered in relatively fixed order, all focused on one disorder or problem or a homogeneous cluster. Such therapies may have only so much capacity for benefit because (a) the narrow problem focus may clash with the comorbidity that is so pervasive in troubled and treated youths (Garber & Weersing, 2010; Kessler, Avenevoli, McLaughlin, et al., 2012; Kuja-Halkola, Lichtenstein, D'Onofrio, & Larsson, 2015; Wolff & Ollendick, 2006), and (b) the standardized sequential designs may clash with the flux in young people's most pressing problems that is so common during episodes of youth psychotherapy (Ollendick & King, 1994; Weisz et al., 2015). Outcomes could conceivably be improved via treatment designs that are more transdiagnostic, flexible, and personalizable (Ng & Weisz, 2016), including perhaps those built from the "elements approach" described by the Institute of Medicine (2015). In that approach, elements of standard ESTs for multiple disorders and problems can be used to form modules (e.g., graduated exposure, cognitive restructuring) and organized into a kind of menu from which personally tailored treatment can be fashioned for each individual and adjusted as treatment needs change. This flexible, personalizable, transdiagnostic approach—built from EST components—has performed well in recent youth RCTs (Chorpita et al., 2017; Weisz et al., 2012).

Overdetermination of outcomes. A third possible explanation for the very limited evidence of improvement over time is that many factors other than psychotherapy may influence outcomes (Institute of Medicine, 2015; Wampold et al., 1997; Weisz et al., 2013), especially for young people. Youths in therapy may experience intrafamily conflict, maltreatment by caregivers, hunger, loss of loved ones, social rejection by peers, academic stress, neighborhood risk, and diverse other forces potentially more powerful than one therapy session per week—in part because youths are essentially confined within family, school, neighborhood, and social systems they cannot escape or avoid and within which their power to exert change is severely limited. Because psychotherapy is but one causal force among many in the lives of young people, there may be a natural upper limit to the impact therapy alone can have within this age range. It is possible that youth treatment developers were actually approaching that upper limit decades ago for some youth problems, with effects of psychotherapy for those problems resting near a natural ceiling ever since. That possibility suggests another strategy for improving outcomes: combining psychotherapy with in vivo support for addressing real-world circumstances that could otherwise limit improvement and over which youths acting alone would have little control. Such an approach would contrast with the primarily office-based approach that has dominated youth psychotherapy research and practice for many years, but a few innovative intervention researchers have achieved success pioneering this more ecologically embedded approach (e.g., Buchanan, Chamberlain, & Smith, 2017; Henggeler & Schaeffer, 2016; Walsh, Theodorakakis, & Backe, 2016).

Change over time in the nature of youth dysfunction and treatment needs. Another factor contributing to our findings may be that the nature of childhood and adolescence, and of youth dysfunction, may be changing faster than our treatments are. Threats to youth mental health are becoming more diverse and multiform than could have been envisioned decades ago. Current threats encompass pressures to excel in increasingly competitive academic and social environments, images conveyed via advertising and social media that could make anyone feel inadequate, risks of harm via text messages and cyberbullying, and even fear of being gunned down at school. These changes may be continually expanding and diversifying the ways youth anxiety and depression are experienced and at a pace well beyond what treatment developers can match. Similarly, the flavor of ADHD—and how it needs to be addressed in treatment—may have been altered significantly by the emerging information age, with television, then the Internet, then video games and smartphones, offering an ever-expanding array of ways to be distracted at the same time as the need for focus and close attention in classroom and social contexts is escalating.

Finally, there are now more ways than ever for youths with conduct problems to be a threat to others than in years past, with available tools that have come to include social media, firearms, and enough information online to turn anyone into a genuine danger—combined with personal access to peer and media influence that can be very difficult for parents to monitor. In sum, social and technological change are continually altering and expanding the range of ways young people may experience anxiety, depression, ADHD, and conduct problems, generating a need for corresponding change in interventions, but at a pace treatment developers may find difficult to match. If treatments for young people are to improve over time, their design and content may need to keep pace with temporal changes in the nature of youth and youth dysfunction, and this is a challenge worthy of our best minds.

Change over time in the culture of parenting, youth communication, and personal change. Societal evolution includes change in parenting standards; increasing print and social media attacks on time-out, for example, may have discouraged the use of methods that have strong empirical support. Social change also includes continual shifts in the ways young people communicate and achieve personal change. Therapies that have worked in the past may need to evolve to synch up with changing patterns of communication and social exchange. For youths accustomed to texting, tweeting, instant messaging, and Snapchat, the idea of sitting alone in a room with a middle-aged adult just talking for 50 min, every week for 20 weeks, may seem like sheer torture—or even a peek into the olden days, like visiting Jurassic Park. Fifty years ago, the creation of therapy manuals that specified details of multiple lengthy sessions in an office with a therapist was a major advance, providing the kind of documentation needed to move psychotherapy beyond unspecified, untested procedures. But times have changed in so many ways since then. If therapy is a form of communication, then its capacity for continual improvement may rest in part on its capacity to evolve continually to fit the communication style of each era. If we are to fit therapies into the increasingly efficient and increasingly electronic communication style of the current era, we are likely to need interventions that can work fast, and many of these may need to rest on digital platforms, live at least partially within computers and smartphone apps, and use strategies that build and sustain youth engagement. Fortunately, these directions are well-represented in recent work, with active discussion and ongoing refinement under way (see, e.g., Fleming et al., 2016; Kaplan & Stone, 2013; Mohr, Weingardt, Reddy, & Schueller, 2017; Schueller, Tomasino, & Mohr, 2017). Evidence on the effects of e-therapies and very brief treatments has been encouraging to date (e.g., Ebert et al., 2015; Schleider & Weisz, 2017, 2018), but whether these emerging approaches can generate improved effects over time will of course not be known for years.

Skeuomorphic thinking in treatment design. In any efforts to synch youth treatment development with changes in the nature of youth psychopathology and in communication trends, a key challenge will be avoiding skeuomorphic thinking. Skeuomorphs are products that retain unnecessary, often ornamental, design features derived from earlier versions of those products. Examples include software calendars that retain the appearance of paper calendars, chandelier light bulbs shaped like candles, e-books with "pages" that appear to turn, and the shutter sound made when we snap a digital photo (Pogue, 2013; Schueller, Munoz, & Mohr, 2013). Experts in the design of digital technologies for mental health intervention (see Schueller et al., 2013) have noted that many efforts to modernize psychotherapies using technology show skeuomorphic thinking that can limit treatment appeal and impact. Examples include e-therapies organized into "sessions," guided by "questionnaires" that look like paper-and-pencil measures and complemented by "workbooks" that look like printed brochures—all unnecessary for effective intervention and potentially counterproductive. Effective treatments may include some that look less like traditional therapy than like video games and other engaging youth pastimes (see, e.g., Fleming, Dixon, Frampton, & Merry, 2012), and some of the best skeuomorph-free ideas may come from young people themselves, if we merely ask them using participatory design approaches (see, e.g., Adkins et al., 2017; Yarosh & Schueller, 2017). If we are to link treatment development to societal change, we may need to remember that old habits die hard and that some of those habits may be counterproductive; user-centered design (Lyon & Koerner, 2016) and shared decision making in treatment planning (Langer & Jensen-Doss, 2018) may help us counter old habits with fresh thinking.

Conclusion

Taken together, our findings highlight the value of examining not only the overall effects of psychotherapy but also the trajectories of change in those effects across years of research. Empirically tested treatments for youth dysfunction represent an enormous scientific advance, with major clinical payoff for young people. However, ideally, the accumulation of treatment development and testing across decades would also lead to increments over time in the magnitude of treatment benefit; we found little evidence of such increments. If the approaches to youth treatment development and testing that have added so richly to intervention science have, in fact, not produced measurable gains over time, fresh ideas may be needed. A useful question for the field is whether there are new approaches that can produce the upward trajectory in treatment benefit toward which so many clinical scientists and clinicians are striving and from which so many troubled youths and families could benefit.

Action Editor

Darby Saxbe served as action editor and June Gruber served as interim editor-in-chief for this article.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The project was supported in part by National Institute of Mental Health Grants MH068806, MH085963, and MH093511 and the Norlien Foundation.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/1745691618805436

References

Adkins, E., Zalta, A. K., Boley, R. A., Glover, A., Karnik, N. S., & Schueller, S. M. (2017). Exploring the potential of technology-based mental health interventions for homeless youth: A qualitative study. *Psychological Services*, *14*, 238–245. doi:10.1037/ser0000120

Agency for Healthcare Research and Quality. (2018). Clinical guidelines and recommendations. Retrieved from http://www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/index.html

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association. (2008). Reporting standards for research in psychology. *American Psychologist*, *63*, 839–851. doi:10.1037/0003-066X.63.9.839

Angold, A., Costello, E. J., Farmer, E. M. Z., Burns, B. J., & Erkanli, A. (1999). Impaired but undiagnosed. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*, 129–137. doi:10.1097/00004583-199902000-00011

Annie E. Casey Foundation. (2018). Evidence2Success. Retrieved from https://www.aecf.org/work/evidence-based-practice/evidence2success/

Buchanan, R., Chamberlain, P., & Smith, D. K. (2017). Treatment Foster Care Oregon: Research and implementation. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (3rd ed., pp. 177–196). New York, NY: Guilford.

California Institute for Behavioral Health Solutions. (2018). Values-driven evidence-based practices. Retrieved from https://www.cibhs.org/evidence-based-practice-initiative

California Social Work Education Center. (2012). Evidence-based practice. Retrieved from https://calswec-archive.berkeley.edu/evidence-based-practice

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52*, 685–716. doi:10.1146/annurev.psych.52.1.685

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., . . . Krull, J. L. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology*, *85*, 13–25. doi:10.1037/ccp0000133

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

Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: The IAPT experience. *International Review of Psychiatry*, *23*, 318–327. doi:10.3109/09540261.2011.606803

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. doi:10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Erlbaum.

Costello, E. J., Angold, A., & Keeler, G. P. (1999). Adolescent outcomes of childhood disorders: The consequences of severity and impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*, 121–128. doi:10.1097/00004583-199902000-00010

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*, 837–844. doi:10.1001/archpsyc.60.8.837

Dorsey, S., McLaughlin, K. A., Kerns, S. E. U., Harrison, J. P., Lambert, H. K., Briggs, E. C., . . . Amaya-Jackson, L. (2016). Evidence base update for psychosocial treatments for children and adolescents exposed to traumatic events. *Journal of Clinical Child & Adolescent Psychology*, *46*, 303–330. doi:10.1080/15374416.2016.1220309

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Ebert, D. D., Zarski, A., Christensen, H., Stikkelbroek, Y., Cuijpers, P., Berking, M., & Riper, H. (2015). Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: A meta-analysis of randomized controlled outcome trials. *PLOS ONE*, *10*(3), Article e0119895. doi:10.1371/journal.pone.0119895

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. doi:10.1136/bmj.315.7109.629

Evans, S. W., Owens, J. S., & Bunford, N. (2014). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, *43*, 527–551. doi:10.1080/15374416.2013.850700

Fleming, T., Dixon, R., Frampton, C., & Merry, S. (2012). A pragmatic randomized controlled trial of computerized CBT (SPARX) for depression among adolescents alienated from mainstream education. *Behavioural and Cognitive Psychotherapy*, *40*, 529–541. doi:10.1017/S1352465811000695

Fleming, T. M., de Beurs, D., Khazaal, Y., Gaggioli, A., Riva, G., Botella, C., . . . Riper, H. (2016). Maximizing the impact of e-therapy and serious gaming: Time for a paradigm shift. *Frontiers in Psychiatry*, *7*, Article 65. doi:10.3389/fpsyt.2016.00215

Fonagy, P., Cottrell, D., Phillips, J., Bevington, D., Glaser, D., & Allison, E. (2015). *What works for whom? A critical review of treatments for children and adolescents*. New York, NY: Guilford.

Freeman, J., Garcia, A., Frank, H., Benito, K., Conelea, C., Walther, M., & Edmunds, J. (2014). Evidence base update for psychosocial treatments for pediatric obsessive-compulsive disorder. *Journal of Clinical Child & Adolescent Psychology*, *43*, 7–26. doi:10.1080/15374416.2013.804386

Garber, J., & Weersing, V. R. (2010). Comorbidity of anxiety and depression in youth: Implications for treatment and prevention. *Clinical Psychology: Science and Practice*, *17*, 293–306. doi:10.1111/j.1468-2850.2010.01221.x

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.

Henggeler, S. W., & Schaeffer, C. (2016). Multisystemic therapy: Clinical overview, outcomes, and implementation research. *Family Process*, *55*, 514–528. doi:10.1111/famp.12232

Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child & Adolescent Psychology*, *45*, 91–113. doi:10.1080/15374416.2015.1046177

Institute of Medicine. (2015). *Psychosocial interventions for mental and substance use disorders: A framework for establishing evidence-based standards*. Washington, DC: National Academies Press. doi:10.17226/19013

Kaminski, J. W., & Claussen, A. H. (2017). Evidence base update for psychosocial treatments for disruptive behaviors in children. *Journal of Clinical Child & Adolescent Psychology*, *46*, 477–499. doi:10.1080/15374416.2017.1310044

Kaplan, R. M., & Stone, A. A. (2013). Bringing the lab and clinic to the community: Mobile technologies for health promotion and disease prevention. *Annual Review of Psychology*, *64*, 471–498. doi:10.1146/annurev-psych-113011-143736

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, *3*, 1–27. doi:10.1146/annurev.clinpsy.3.022806.091432

Kessler, R. C., Avenevoli, S., Costello, E. J., Georgiades, K., Green, J. G., Gruber, M. J., . . . Merikangas, K. R. (2012). Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the national comorbidity survey replication adolescent supplement. *Archives of General Psychiatry*, *69*, 372–380. doi:10.1001/archgenpsychiatry.2011.160

Kessler, R. C., Avenevoli, S., McLaughlin, K. A., Green, J. G., Lakoma, M. D., Petukhova, M., . . . Merikangas, K. R. (2012). Lifetime co-morbidity of DSM-IV disorders in the US national comorbidity survey replication adolescent supplement (NCS-A). *Psychological Medicine*, *42*, 1997–2010. doi:10.1017/S0033291712000025

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, *62*, 593–602. doi:10.1001/archpsyc.62.6.593

Kuja-Halkola, R., Lichtenstein, P., D'Onofrio, B. M., & Larsson, H. (2015). Codevelopment of ADHD and externalizing behavior from childhood to adulthood. *Journal of Child Psychology and Psychiatry*, *56*, 640–647. doi:10.1111/jcpp.12340

Langer, D. A., & Jensen-Doss, A. (2018). Shared decision-making in youth mental health care: Using the evidence to plan treatments collaboratively. *Journal of Clinical Child and Adolescent Psychology*, *47*, 821–831. doi:10.1080/15374416.2016.1247358

Lipsey, M., & Wilson, D. (2000). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.

Lyon, A. R., & Koerner, K. (2016). User-centered design for psychosocial intervention development and implementation. *Clinical Psychology: Science and Practice*, *23*, 180–200. doi:10.1111/cpsp.12154

McCart, M. R., & Sheidow, A. J. (2016). Evidence-based psychosocial treatments for adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology*, *45*, 529–563. doi:10.1080/15374416.2016.1146990

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*, 361–365. doi:10.1037/0033-2909.111.2.361

Merikangas, K., Hep, J., Burstein, M., Swanson, S., Avenevoli, S., Cui, L., . . . Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey-Replication Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*, 980–989. doi:10.1016/j.jaac.2010.05.017

Miller, S., Wampold, B., & Varhely, K. (2008). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*, *18*, 5–14. doi:10.1080/10503300701472131

Mohr, D. C., Weingardt, K. R., Reddy, M., & Schueller, S. M. (2017). Three problems with current digital mental health research . . . and three things we can do about it. *Psychiatric Services in Advance*, *5*, 427–429. doi:10.1176/appi.ps.201600541

Nathan, P. E., & Gorman, J. M. (2015). *A guide to treatments that work* (4th ed.). New York, NY: Oxford University Press.

National Institute of Mental Health. (2016, September 8). *Psychosocial research at NIMH: A primer*. Retrieved from https://www.nimh.nih.gov/research-priorities/psychosocial-research-at-nimh-a-primer.shtml#2

Ng, M. Y., & Weisz, J. R. (2016). Building a science of personalized intervention for youth mental health. *Journal of Child Psychology and Psychiatry*, *57*, 216–236. doi:10.1111/jcpp.12470

Ollendick, T. H., & King, N. J. (1994). Diagnosis, assessment, and treatment of internalizing problems in children. *Journal of Consulting and Clinical Psychology*, *62*, 918–927. doi:10.1037/0022-006X.62.5.918

Pogue, D. (2013). Out with the real. *Scientific American*, *308*(2), 29. doi:10.1038/scientificamerican0213-29

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: SAGE.

Saletta, F., Seng, M. S., & Lau, L. M. S. (2014). Advances in paediatric cancer treatment. *Translational Paediatrics*, *3*, 156–182. doi:10.3978/j.issn.2224-4336.2014.02.01

Schleider, J. L., & Weisz, J. R. (2017). Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*, 107–115. doi:10.1016/j.jaac.2016.11.007

Schleider, J. L., & Weisz, J. R. (2018). A single-session growth mindset intervention for adolescent anxiety and depression: Nine-month outcomes of a randomized trial. *Journal of Child Psychology and Psychiatry*, *59*, 160–170. doi:10.1111/jcpp.12811

Schueller, S. M., Munoz, B. F., & Mohr, D. C. (2013). Realizing the potential of behavioral intervention technologies. *Current Directions in Psychological Science*, *22*, 478–483. doi:10.1177/0963721413495872

Schueller, S. M., Tomasino, K. N., & Mohr, D. C. (2017). Integrating human support into behavioral intervention technologies: The efficiency model of support. *Clinical Psychology: Science and Practice*, *24*, 27–45. doi:10.1111/cpsp.12173

Silverman, W. K., & Hinshaw, S. P. (Eds.). (2008). Evidence-based psychosocial treatments for children and adolescents: A ten year update [Special issue]. *Journal of Clinical Child & Adolescent Psychology*, *37*(1). doi:10.1080/15374410701817725

Smith, M. L. (1980). *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins Press.

Southam-Gerow, M. A., & Prinstein, M. J. (2014). Evidence base updates: The evolution of the evaluation of psychological treatments for children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *43*, 1–6. doi:10.1080/15374416.2013.855128

Substance Abuse and Mental Health Services Administration. (2018). Evidence-based practices resource center. Retrieved from https://www.samhsa.gov/ebp-resource-center

UK National Institute for Health and Care Excellence. (2018). Improving health and social care through evidence-based guidance. Retrieved from https://www.nice.org.uk/

Walsh, M. E., Theodorakakis, M. D., & Backe, S. (2016). Redesigning a core function of schools: A systematic, evidence-based approach to student support. In H. A. Lawson & D. Van Veen (Eds.), *Developing community schools, community learning centers, extended-service schools and multi-service schools* (pp. 127–147). The Hague, The Netherlands: Springer.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, “all must have prizes.” *Psychological Bulletin*, *122*, 203–215. doi:10.1037/0033-2909.122.3.203

Weersing, V. R., Jeffreys, M., Do, M. T., Schwartz, K. T. G., & Bolano, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child & Adolescent Psychology*, *46*, 11–43. doi:10.1080/15374416.2016.1220310

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral and nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*, 317–320. doi:10.1037/0022-006X.63.2.317

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. Cambridge, England: Cambridge University Press.

Weisz, J. R., Chorpita, B. F., Palinkas, L. A., Schoenwald, S. K., Miranda, J., Bearman, S. K., . . . Gibbons, R. D., and the Research Network on Youth Mental Health. (2012). Testing standard and modular designs for psychotherapy with youth depression, anxiety, and conduct problems: A randomized effectiveness trial. *Archives of General Psychiatry*, *69*, 274–282. doi:10.1001/archgenpsychiatry.2011.147

Weisz, J. R., Doss, A. J., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: A meta-analysis of direct comparisons. *American Psychologist*, *61*, 671–689. doi:101037/0003066X617671

Weisz, J. R., & Kazdin, A. E. (Eds.). (2017). *Evidence-based psychotherapies for children and adolescents* (3rd ed.). New York, NY: Guilford.

Weisz, J. R., Krumholz, L. S., Santucci, L., Thomassin, K., & Ng, M. (2015). Shrinking the gap between research and practice:

Tailoring and testing youth psychotherapies in clinical care contexts. *Annual Review of Clinical Psychology*, *11*, 139–163. doi:10.1146/annurev-clinpsy-032814-112820

Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA Psychiatry*, *70*, 750–761. doi:10.1001/jamapsychiatry.2013.1176

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., . . . Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *American Psychologist*, *72*, 79–117. doi:10.1037/a0040360

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth psychotherapy in the mental health ecosystem. *Journal of Clinical Child and Adolescent Psychology*, *42*, 274–286. doi:10.1080/15374416.2013.764824

Weisz, J. R., Weiss, B., Han, 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. doi/10.1037/0033-2909.117.3.450

Wolff, J. C., & Ollendick, T. H. (2006). The comorbidity of conduct problems and depression in childhood and adolescence. *Clinical Child and Family Psychology Review*, *9*, 201–220. doi:10.1007/s10567-006-0011-3

Yarosh, S., & Schueller, S. M. (2017). “Happiness inventors”: Informing positive computing through participatory design with children. *Journal of Medical Internet Research*, *19*(1), Article e14. doi:10.2196/jmir.6822