


Two-Year Impact of Prevention Programs on Adolescent Depression: an Integrative Data Analysis Approach

C. Hendricks Brown¹  · Ahnalee Brincks^{2,3} · Shi Huang⁴ · Tatiana Perrino³ · Gracelyn Cruden^{1,5} · Hilda Pantin³ · George Howe⁶ · Jami F. Young⁷ · William Beardslee⁸ · Samantha Montag¹ · Irwin Sandler⁹

Published online: 24 December 2016
© Society for Prevention Research 2016

Abstract This paper presents the first findings of an integrative data analysis of individual-level data from 19 adolescent depression prevention trials ($n = 5210$) involving nine distinct interventions across 2 years post-randomization. In separate papers, several interventions have been found to decrease the risk of depressive disorders or elevated depressive/internalizing symptoms among youth. One type of intervention specifically targets youth without a depressive disorder who are at risk due to elevated depressive symptoms and/or having a parent with a depressive disorder. A second type of intervention targets two broad domains: prevention of problem behaviors, which we define as drug use/abuse, sexual risk behaviors, conduct disorder, or other externalizing problems, and general mental health. Most of these latter interventions improve parenting or family factors. We examined the shared and unique effects of these interventions by level of baseline youth depressive symptoms, sociodemographic

characteristics of the youth (age, sex, parent education, and family income), type of intervention, and mode of intervention delivery to the youth, parent(s), or both. We harmonized eight different measures of depression utilized across these trials and used growth models to evaluate intervention impact over 2 years. We found a significant overall effect of these interventions on reducing depressive symptoms over 2 years and a stronger impact among those interventions that targeted depression specifically rather than problem behaviors or general mental health, especially when baseline symptoms were high. Implications for improving population-level impact are discussed.

Keywords Data synthesis · Person-level meta-analysis · Growth modeling · Growth mixture modeling · Parenting interventions · Cognitive behavioral therapy · Interpersonal therapy

Electronic supplementary material The online version of this article (doi:10.1007/s11121-016-0737-1) contains supplementary material, which is available to authorized users.

✉ C. Hendricks Brown
hendricks.brown@northwestern.edu

¹ Departments of Psychiatry and Behavioral Sciences, Preventive Medicine, and Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

² Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA

³ Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, USA

⁴ Department of Biostatistics, Vanderbilt University, Nashville, TN, USA

⁵ Department of Health Policy and Management, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

⁶ Departments of Psychology and Psychiatry and Behavioral Sciences, George Washington University, Washington, DC, USA

⁷ Graduate School of Applied and Professional Psychology, Rutgers University, New Brunswick, NJ, USA

⁸ Boston Children's Hospital, Harvard University, Cambridge, MA, USA

⁹ Department of Psychology, Arizona State University, Tempe, AZ, USA

Introduction

The prevention of depressive disorders and subclinical depressive symptoms is an important public health priority (Bertha and Balázs 2013; Institute of Medicine and National Research Council 2009; Munoz et al. 2012; Wesselhoeft et al. 2013). Research reviews and meta-analyses of trials testing existing interventions for youth suggest that depressive symptoms can be reduced and depressive disorders prevented, at least in the short term (Horowitz and Garber 2006; Institute of Medicine and National Research Council 2009; Merry et al. 2011; Munoz et al. 2012; Stice et al. 2009). Yet, these interventions vary greatly in terms of the form of intervention they employ, the types of youth they target, and the putative mechanisms through which these interventions are hypothesized to work. A recent overview of meta-analyses for depression in youth with significant heterogeneity across trials (Sandler et al. 2014), although relatively little information is available to account for such variation in impact. For preventive interventions to be maximally effective, it is important to understand “for whom, how long, and under what circumstances” these interventions work (Brown et al. 2009; Stice et al. 2009).

Some researchers have tested whether intervention effects vary across sample characteristics such as initial symptom severity (Sandler et al. 2003), gender (Clarke et al. 2001), or race and ethnicity (Cardemil et al. 2007; Marchand et al. 2010), but findings have not been consistent, and individual trials are often underpowered for detecting such heterogeneity (Brown et al. 2013). Meta-analyses using meta-regression provide some clues that preventive effects may vary depending on age, gender, and initial symptom severity (Horowitz and Garber 2006; Merry et al. 2011; Stice et al. 2009), but standard meta-regression methods, which depend on effect sizes aggregated at the trial level, are severely underpowered for detecting such moderation effects (Dagne et al. 2016) and may be so severely biased that conclusions from aggregate data may be in the opposite direction of the true moderated effects at the individual level (Petkova et al. 2013).

In this paper, we report results based on an aggregated dataset that combines longitudinal data from 5210 adolescents participating in 19 trials of interventions that were either designed to prevent depression or to prevent a broader set of behavioral problems or improve mental health. The use of individual-level data in this large dataset substantially increases power to detect heterogeneity while avoiding many of the pitfalls of standard meta-regression techniques (Brown et al. 2013). These data involve individual-level measurement of depressive symptoms through 2 years post-randomization, baseline participant characteristics, as well as intervention attributes that potentially serve as moderators of intervention effects. We employ novel statistical methods,

combining techniques from integrative data analysis (Curran and Hussong 2009; Greenbaum et al. 2015), latent variable analysis, and growth modeling (see Brincks et al. 2016) to test a series of hypotheses concerning heterogeneity of prevention effects.

Potential Sources of Heterogeneity of Intervention Effects

Existing prevention programs have been based on several different theoretical models. These include cognitive-behavioral theory (CBT), which focuses on improving youth cognitive styles and building coping skills (e.g., Clarke et al. 2001; Garber et al. 2009; Gillham et al. 2012; Horowitz et al. 2007); interpersonal psychotherapy (IPT), which aims to improve interpersonal skills and relationships and reduce conflict (e.g., Horowitz et al. 2007; Young et al. 2006; Young et al. 2010); and parenting and family systems (PFS) interventions, which attempt to improve parenting practices and family relationships (e.g., Connell and Dishion 2008; Gonzales et al. 2012; Prado and Pantin 2011; Trudeau et al. 2012). Some programs also combine approaches such as parenting and coping training (e.g., Compas et al. 2009; Sandler et al. 2003; Wolchik et al. 2000). Finally, some trials have examined the potential incremental benefits of adding a parent component to CBT- or IPT-based child-centered, preventive interventions and have found no differences in effects (e.g., Gillham et al. 2012; Young et al. 2010), though the studies may have been underpowered to detect such effects.

Although the theoretical orientation of the intervention may influence intervention response, few studies have compared the effects of different types of preventive interventions. Horowitz et al. (2007) compared IPT and CBT prevention programs and found that, by the end of the intervention, both reduced depressive symptoms with no significant differences between them, although both were significantly better than the no-intervention control condition. On the other hand, a meta-analysis of preventive interventions across the life span found some evidence that IPT-based programs were more effective than CBT-based interventions, but cautioned that this was based on only a few IPT-based studies (Cuijpers et al. 2008). We include trials testing all three types of interventions (CBT, IPT, and PFS) in the current study, allowing us to test whether intervention type does moderate impact on depression.

Trials have also varied in their primary or focal targets. Some interventions have specifically targeted the prevention of depressive disorders and symptoms (e.g., Clarke et al. 2001; Garber et al. 2009; Gillham et al. 2012; Horowitz et al. 2007), whereas others have focused more broadly on promoting general mental health and preventing both internalizing and externalizing conditions (e.g., Sandler et al. 2003; Wolchik et al. 2000). Others have aimed to prevent youth externalizing symptoms (e.g., drug abuse and conduct problems), yet have found significant protective effects on youth

depressive and internalizing symptoms as well (Connell and Dishion 2008; Perrino et al. 2014; Trudeau et al. 2012). These secondary effects on internalizing symptoms may be due to the influence of these interventions on common protective processes, such as positive parenting and family relationships (Perrino et al. 2014; Restifo and Bogels 2009; Sander and McCarty 2005), or to longer-term developmental changes passing first through reductions in externalizing behavior (McClain et al. 2010; Perrino et al. 2016). To date, no meta-analysis of the prevention of depression in youth has included intervention trials targeting externalizing problems, risky sexual behavior, or drug abuse. Thus, it is unclear whether the effects of these programs are similar to depression-specific programs and how the inclusion of these trials impacts overall effect sizes. To test this question, we included trials with different focal targets (depression alone, problem behavior, and general mental health), allowing us to test whether the targets of these trials had different impacts on preventing depression.

There is also evidence from individual trials that population characteristics contribute to heterogeneity in the effects of preventive interventions on depressive disorders and depressive and internalizing symptoms. Individual trials have found that adolescent subgroups show different intervention responses depending on age, gender, ethnicity, and initial levels of depressive symptoms (Horowitz and Garber 2006; Merry et al. 2011; Stice et al. 2009). The meta-analysis of Stice et al. (2009) of trials targeting depression found greater trial-level effects for interventions involving older, higher-risk participants (e.g., those with elevated depressive symptoms), as well as samples with greater proportions of female and non-white participants. Horowitz and Garber's (2006) earlier meta-analysis of similar trials also concluded that there were larger effects for trials with higher proportions of females and higher-risk participants and identified larger effects for older versus younger adolescent participants.

However, findings from some individual trials have not been consistent with the meta-analyses previously mentioned. For instance, Clarke et al. (1993) found that boys responded more positively to a CBT-based preventive intervention than did girls. Regarding ethnicity, the Penn Resilience Program found beneficial effects on depressive symptoms for Hispanic but not African-American youth (Cardemil et al. 2007). Using data from two studies, Marchand et al. (2010) found no differential effects of a CBT-based prevention program for Asian, Latino, and European American adolescents. Because many trials had limited representation of the broad diversity of youth, further research is needed to determine whether and which demographic or clinical variables moderate intervention outcomes. In the current study, we tested whether key demographic variables, including gender, age, race and ethnicity, parent education, and family income, as well as initial symptom severity moderated the impact of interventions on preventing depressive symptoms.

Analysis of Summary-Level versus Participant-Level Data across Trials

The inconsistent findings on moderators of intervention effects, together with the substantial variability across interventions and their samples, suggest the need for more in-depth approaches to synthesis of data from multiple trials. Single randomized trials are often underpowered to detect intervention moderator effects and do not consistently report tests of the same moderator variables (Brown et al. 2013). Although meta-analysis has important strengths for the examination of main effects across trials, its utility for understanding intervention moderator effects is more limited (Brown et al. 2013; Kraemer et al. 2002; Lipsey 2003). One concern is that non-significant moderator analyses typically are not published, limiting the information available to understand moderators through combining results of moderator analyses through meta-analysis. A related concern is that individual trials can conduct moderator analyses using different methods (e.g., continuous versus cut points) and different variables, making it difficult to draw clear conclusions regarding moderators using a traditional meta-analytic approach (Brown et al. 2013).

A promising strategy for addressing the limitations of single-trial analyses and meta-analyses is to combine datasets from individual trials and conduct integrative data analysis (IDA), also known as individual patient-level meta-analysis (Higgins et al. 2001; Stewart and Clarke 1995). IDA is defined as an "analysis of multiple data sets that have been pooled into one" (Curran and Hussong 2009; Curran et al. 2014; Hussong et al. 2013), and in our case involves pooling individual-level data together across multiple trials and conducting synthesis analyses across similar studies. In general, IDA yields larger sample sizes, increased statistical power, increased variability on important measures, and the capacity to test more sophisticated models. IDA also can be applied to multiple randomized trials to assess moderation and mediation (Brown et al. 2013; Gibbons et al. 2012a, b; Perrino et al. 2014).

IDA is particularly promising for evaluating individual-level moderators. Dagne et al. (2016) found that power to detect moderator effects for individual-level moderators could be as much as 16 times greater for IDA as compared to standard meta-regression. Unlike meta-regression, IDA with individual-level data will not result in the ecological fallacy of interpreting group-level associations as reflecting individual-level effects.

Recent synthesis analyses demonstrate the potential benefits of IDA for understanding intervention moderator effects. For instance, Perrino et al. (2014) combined participant-level data from three trials of the Familias Unidas intervention, a family-focused intervention that has been found to promote family functioning and to prevent drug abuse, externalizing problems, and HIV risk among Hispanic youth (Prado and Pantin 2011). Using IDA techniques, this study found that

parent–adolescent communication, a modifiable risk factor and hypothesized mechanism by which this intervention works, was a significant moderator of intervention effects on internalizing symptoms. Importantly, communication was not a significant moderator of intervention effects when combined analyses were conducted on each separate trial's moderation findings, suggesting that IDA increased the statistical power needed to detect this effect (Perrino et al. 2014).

Similar IDA analyses have been used to examine treatment effects. Gibbons et al. (2012a, b) combined individual-level data from 41 antidepressant trials involving pediatric, adult, and geriatric populations and found that age moderated treatment effects, with greater reductions in depressive symptoms for children and adults as compared to patients over 60 years of age. Greenbaum et al. (2015) used IDA to combine data from trials of five drug abuse treatment programs involving multidimensional family therapy, finding stronger preventive effects on drug abuse for female adolescents and effects only for European American and African-American but not Hispanic participants.

The Current Study

The current paper used IDA to test a series of hypotheses concerning the overall impact of prevention programs over time and whether study-level and individual-level variables moderate the impact of preventive interventions on adolescent depression. We combined participant-level data from 19 randomized controlled trials that tested whether interventions prevented depressive symptoms. The combined dataset includes nine distinct preventive interventions and a total sample of 5210 adolescent participants measured up to six times over 2 years after baseline, depending on the trial protocol. IDA was used to examine intervention effects on adolescent depressive symptom outcomes, as well as to identify factors that moderated intervention effects. Specifically, this paper examined the following questions: (1) Do preventive interventions have an overall effect on the course of adolescent depressive symptoms? (2) If so, does this effect remain constant (linear), increase, or decrease over time? (3) Do intervention effects differ by adolescents' gender, age, and baseline level of depressive symptoms? (4) Do effects differ by type of intervention (i.e., CBT, IPT, and PFS)? (5) Do effects differ by focal target of the intervention (i.e., depression alone, problem behaviors, or general mental health)?

Methods

Sample

We procured 19 full datasets from published trials of relevant preventive interventions. These trials all satisfied the

following criteria: (1) the study must be described as having focal targets of adolescent depression alone, problem behaviors, or general mental health; (2) the study must randomize to intervention or control, although randomization could occur at the individual, family, or community level; (3) samples must include target participants between the ages of 11 and 18 upon enrollment, although a study could include other participants outside that age range; (4) the study must include measures of depressive or internalizing symptoms; and (5) symptom measures must be administered in at least one follow-up occurring at least 6 months after baseline, although most had much longer follow-ups.

In obtaining these data from 19 trials, we initially contacted investigators who had conducted 24 prevention trials meeting these criteria and who indicated interest in collaborating. After securing grant funding, we returned to these investigators in order to develop collaborative working arrangements (Perrino et al. 2013) for a detailed discussion of this stage of collaborative engagement. We were ultimately able to gain access to complete individual-level datasets from 18 of these 24 trials and then obtained an additional trial from one of our collaborators, bringing the total to 19 (listed in Table 1 and the PRISMA diagram shown in Fig. 1).

All trials were designed to randomize participants at the individual or family level, with one exception (preparing for the drug-free years, PDFY) which was randomized at the community level. All interventions were delivered by a trained professional and included a range of intervention types such as parent skills training or prevention versions of CBT-, IPT-, or family-based therapy. Whereas some trials targeted the prevention of adolescent depression specifically, others targeted problem behavior or general mental health promotion. Many of the trials included are considered effectiveness trials, although some are in the efficacy evaluation stage.

Some trials included children who were both inside and outside of our targeted age range. In those cases, we selected only participants who fit our age criteria (between 11 and 18), eliminating 248 participants from further analysis. This left a combined sample of 5210. Sample characteristics for each trial are listed in Table 1. Participants ranged in age from preadolescent (ages 11–12: $n = 2751$, 52.9 %) to early adolescent (ages 13–15: $n = 2133$, 40.9 %) to older adolescent (ages 16–18: $n = 326$, 6.3 %). The sample was 49.8 % female; 32.7 % of participants identified as culturally Hispanic, 58.7 % white, 9.5 % African-American, 1.8 % Asian, 2.3 % Native American, 0.6 % Hawaiian or Pacific Islander, 3.8 % other or multiracial, and 23.3 % unknown; most of these unknowns were located in predominantly white, non-Hispanic communities.

Although meta-analytic studies attempt to gather data from all relevant studies ever conducted on a particular topic, this is a difficult goal for IDA where full datasets are required. In order to provide information on how representative these 19

Table 1 Characteristics of the 19 trials

Trial	<i>N</i> original publication	<i>N</i> in our data	Latest follow-up point available to us (months)	No. of intervention arms	Type: IPT, CBT, Family parenting	Target: Parent, child, conjoint	Focus: depression or other problem behavior or general mental health	Citation
ADEPT	94	94	24	2	CBT	P, C	Depression	Clarke AGP 2001
CATCH-IT	83	50	30	2	CBT, IPT	P, C	Depression	Van Voorhees et al. 2009
Family Talk	121	72	48	2	Family	P, C, CJ	Depression	Beardslee et al. 2003
IPT-AST versus CBT	380	379	18	3	CBT, IPT	C	Depression	Horowitz et al. 2007
K-IPT-AST	57	57	18	3	IPT	C, CJ	Depression	Young et al. 2010
NARSAD IPT	41	41	6	2	IPT	C	Depression	Young et al. 2006
Penn Resiliency Program (PRP) I	697	690	36	3	CBT	C	Depression	Gillham et al. 2007
Penn Resiliency Program (PRP) II	408	407	24	3	CBT, Family	P, C	Depression	Gillham et al. 2012
Prevention of Depression Study (PODS)	316	316	6	2	CBT	P, C	Depression	Garber et al. 2009
Bridges (Puentes)	516	542	14	2	Family	P, C, CJ	Other	Gonzales et al. 2012
Familias Unidas I	266	257	36	3	Family	P, CJ	Other	Prado et al. 2007
Familias Unidas II	213	213	30	2	Family	P, CJ	Other	Pantun et al. 2009
Familias Unidas DJJ	242	242	12	2	Family	P, CJ	Other	Prado et al. 2011
Familias Unidas CDC	n/a	160	12	2	Family	P, CJ	Other	Not yet published
Family Bereavement	244	141	72	2	CBT, Family	P, C, CJ	Other	Sandler et al. 2003
New Beginnings	240	99	72	3	CBT, Family	P, C, CJ	Other	Wolchik et al. 2000
Preparing for the drug-free years (PDFY) and Iowa Strengthening Families Program (ISF) Project Alliance I	667	659	80	3	Family	P, C, CJ	Other	Spoth et al. 1995; Spoth et al. 2001
Project Alliance II	672 (998)	178	72	2	Family	C, CJ	Other	Dishion et al. 2002; Connell et al. 2007
Project Alliance II	593	581	60	2	Family	P, C, CJ	Other	Stormshak et al. 2011

P parent, *C* child, *Cj* conjoint

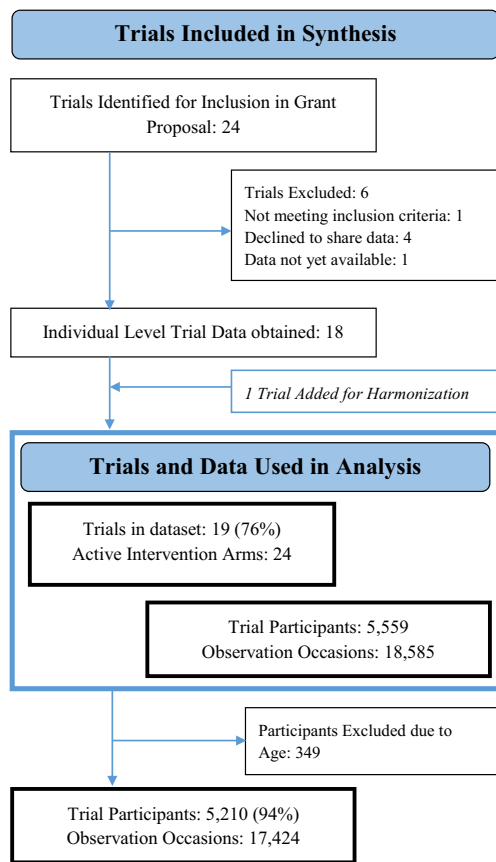


Fig. 1 PRISMA statement

trials might be of the universe of possible trials, we conducted a comprehensive survey of published research using several citation databases. Details of this survey are described in Supplemental Fig. A, and an additional comparison can be found in a companion paper in this issue (Brincks et al. 2016). We identified a total of 68 additional prevention trials meeting our selection criteria. Supplemental Table B compares the 19 trials in our dataset with the other larger set of trials on several dimensions. The primary difference was that many of these other trials were conducted outside the USA. Nine of our 19 trials (47 %) focused on preventing depression or internalizing behavior, whereas 18 (26 %) of the other trials used were focused on preventing depression (Yates corrected $p = 0.14$). We thus obtained 33 % of such existing trials for this synthesis project. Trial sample sizes varied greatly, but had similar means.

Measures

Outcome Measures

As the outcome measures for adolescent depressive and internalizing symptoms varied across the 19 trials, this presented a challenge in harmonizing outcomes across the studies. We identified eight measures of depressive symptoms that were

used across the 19 trials. Some of these were limited to depressive symptoms; others included depressive symptoms mixed with other internalizing symptoms such as anxiety or withdrawal. We decided to keep these latter measures intact rather than using only depressive symptom items, given that they represent established measures of child functioning. In this paper, we refer to this entire set as measures of depressive symptoms, given that depression items make up the vast majority of items.

The eight measures included: (1) Children’s Depression Rating Scale (CDRS-R, clinician-rated; Poznanski and Mokros 1996); (2) Children’s Depression Inventory (CDI, self-report; Kovacs and Beck 1977); (3) Anxiety/Depression subscale of the Youth Self-Report (YSR-ANX/DEP, self-report; Achenbach and Rescorla 2001); (4) Withdrawal/Depression subscale of the Youth Self-Report (YSR-WIT/DEP, self-report; Achenbach and Rescorla 2001); (5) Center for Epidemiologic Studies—Depression Scale (CESD, self-report; Radloff 1977); (6) Anxiety/Withdrawal subscale of the Revised Behavior Problem Checklist (RBPC-ANX/WIT, parent report; Quay 1983); (7) Anxious/Depressed subscale of the Child Behavior Checklist (CBCL-ANX/DEP, parent report; Achenbach 1991); and (8) Withdrawal/Depression subscale of the Child Behavior Checklist (CBCL-WIT/DEP, parent report; Achenbach 1991).

Whenever possible, we scored measures based on original items rather than relying on pre-computed constructs in each of the trial datasets. This allowed us to standardize scoring across all trials (e.g., treat a scale as missing if less than 80 % of its items were completed by the individual).

Some trials did not measure the entire set of available items on a particular instrument. For example, one trial administered a custom set of items closely related to the CDI to all participants and the CDI to a smaller set (23 %). The overlap of items between this custom measure and the CDI allowed us to regress the summary score for the custom measure on CDI in our model and thereby infer CDI scores on all participants. We used a similar approach for one trial that used a shortened version of the CESD at follow-up. With parent-reported instruments, we chose the mother’s report over the father’s when available because the correlations between father’s report and other measures were not as strong as those for the mother. In the few cases where the mother’s report was not available, we used the father’s report.

Candidate Moderators

We limited our candidate moderators to variables that were collected in all or nearly all of our 19 trials. This included individual-level sociodemographic characteristics including race/ethnicity, age, gender, and family socioeconomic status, baseline level of depressive symptoms, characteristics of the intervention, and how it was delivered.

Race/ethnicity was coded into the following non-mutually exclusive categories: Hispanic, Black/African-American, Native American, Asian, White, Hawaiian, or Other/Multiple Race. When Hispanic ethnicity was missing or unknown, participants were categorized based on their self-identified race. Whereas the U.S. Census and the National Institutes of Health now consider race and ethnicity to be different concepts for which data should be collected separately (Grieco and Cassidy 2001; National Institutes of Health 2001), not all studies collected both race and ethnicity data, resulting in some incomplete data on these potential modifiers. Although there are limitations to the race/ethnicity categorization utilized in some of the earlier trials in this study, we decided to categorize Hispanics as a separate category that collapses across racial groups based on existing evidence that a significant proportion of Hispanics self-identify as “Hispanic” when asked to select a racial category, suggesting that this may better reflect the perceived identities of many U.S. Hispanic individuals (Hitlin et al. 2006). This race/ethnicity categorization also maximizes the use of available data as some studies with Hispanics did not ask about race. For analyses where numbers were too small to conduct separate race/ethnicity analyses, we created a minority classification by including all groups other than white non-Hispanic. A separate binary variable coded participants as minority (i.e., any racial or ethnic minority group) versus non-minority (i.e., white).

Other sociodemographic variables considered as moderators included age, gender, parent education, and family income, all measured by parent report. Parent education was indexed as the highest grade attained by either parent. Annual family income in dollars was measured by parent report, using the mother’s report in the few instances that the two parents’ reports differed. In trials where income was reported by selecting one of several range categories, we used the midpoint of the range category endorsed.

Trial- and arm-level variables included intervention type, target participant, and focal target. Using published reports of each trial, three staff members coded each trial independently on these variables, and we sought confirmation of the final coding from the primary study authors. Six trials included more than one active intervention arm, so several variables were coded by intervention arm rather than by trial. *Intervention type* was coded as one of three categories: cognitive behavioral, interpersonal, or family-based/parenting. Target participant variables indicated to whom the intervention was delivered. Three separate variables were developed to account for interventions with multiple components. *Delivered to child* was coded as a yes/no variable, indicating whether the child received at least a portion of the intervention without the parent present, in a youth group or individually. *Delivered to parent* was categorized to distinguish whether and how the parent received the intervention. This variable

was coded using three possible codes: (1) parent received information only about what their child was learning in the intervention; (2) parent received more formal therapeutic intervention than information (e.g., cognitive behavioral or parenting skills training); or (3) no parent contact. *Focal target* was coded as a binary variable: we identified trials as depression-focused when the explicitly stated target of the intervention was the prevention of a depressive disorder or reduction of depressive symptoms and as problem behavior or general mental health when the focal target was some form of problem behavior or broader mental health promotion.

Missing Data Strategy

The dataset formed by combining data from the 19 trials was quite complex because each trial employed only a small subset of the eight depressive symptom measures and different trials used different follow-up periods. We approached this as a multivariate missing data problem. The details of this modeling approach are reported in Brincks et al. (2016); here, we provide a brief overview.

Scores on depression measures were not available when participants failed to complete a measure or participate in a particular measurement occasion. These scores were all set to missing. In addition, when a trial did not use a particular depression measure, we assigned scores on that measure to missing for all participants in that trial. We used standard full-information maximum likelihood methods to deal with these forms of missingness. These methods assume that data are missing at random; that is, the occurrence of missing data is not associated with unobserved values (Rubin 1976). . At the trial level, missingness often occurred because a measure was not used in that trial. By conditioning on trial in our analysis, this is a type of missingness at random. Within each trial, missingness was primarily due to attrition. Because each trial had follow-up evaluations conducted by assessors masked to the intervention condition, any potential departures from missing at random within a trial would likely be comparable across conditions.

Given that trials used different follow-up periods, we attempted to employ random effects modeling methods that allowed time of measurement to vary across respondents. These models failed to converge, given the complexity of the full model. We then decided to use a harmonization strategy.

In order to use general models involving multivariate growth modeling for panel data, we condensed follow-up measures into six time blocks that had reasonably narrow widths. We did this because the analysis of panel data provided more flexibility than did other procedures that treated follow-up times as continuous (e.g., we could examine growth mixture models; Muthén et al. 2002). The resulting time blocks were: baseline (time 0); 7 days to 2 months (time 1);

2–3 months (time 2); 5.5–6 months (time 3); 9–14 months (time 4); 14–18 months (time 5); and 24 months (time 6). Figure 2 identifies the number of youth at baseline enrolled into the trial within ages 11–18 and shows the availability of follow-up data for each of the trials at each time period up to 24 months post-intervention. Note that all time blocks had some data available across all these 19 trials, with five of the trials measured up to 24 months, and all trials had at least one assessment through 14–18 months. Only a few trials had measures at time 1 (7 days to 2 months), and all of those that did also had measures at time 2 (2–3 months). Because of the limited amount of data available at the first follow-up time block (time 1), the estimates at this time point were expected to be unstable. As all of the trials that had data at time 1 also had data at time 2, we did not include the first follow-up data in our modeling. This resulted in the loss of only 2.9 % of all data.

Statistical Modeling

Testing for Change across Time

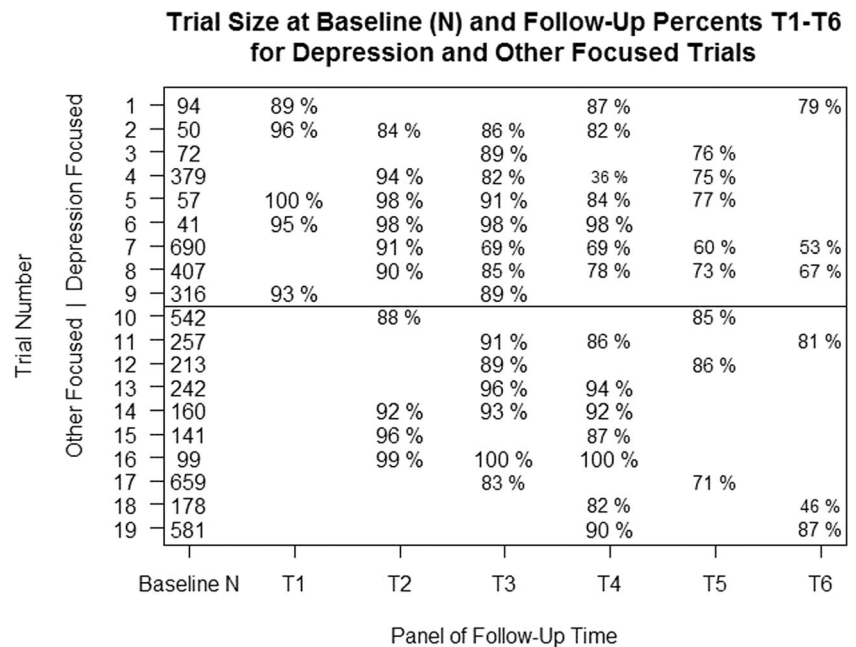
The major analytic challenge in this paper summarizes individual growth trajectories across time and outcome measure. At each time block, there were multiple measures of depressive symptoms, with varying levels of overlap across trials, as specified in Siddique et al. (2016). Because all the depressive symptom measures related to a single underlying construct, we chose to model these multiple measures as a latent variable, in a second-order latent growth model, as described in detail in Brincks et al. (2016). Across the six time blocks, we constrained the loadings and intercepts on the measurement

model for depressive symptoms to be equal and allowed the factor variances to change with time and covariates. This allowed us to examine how trajectories of these latent variables changed over time.

We allowed our modeling of the growth over time, i.e., “slope,” to capture any linear as well as nonlinear pattern across time points. This is especially important in prevention as many prevention effects may diminish or even reverse over time, and such patterns would not be detected if we forced the pattern to be linear. Nonlinearity was accounted for by allowing the second-level factor loadings of the growth model to be estimated by the data, allowing the estimation of all but the first two loadings, which were fixed to force identifiability.

Our second-order latent growth model included a latent variable for baseline depressive symptoms (Intercept) and a latent variable for change on depressive symptoms (Slope); see equations in Brincks et al. (2016). In our structural equation modeling, we controlled on demographic variables such as age, gender, race/ethnicity, family income, and parent’s educational attainment for both the second-level intercept and slope. We also adjusted the intercept and the slope, for trial as a fixed effect in all these second-order latent growth models. This introduction of fixed effects for trial was used instead of random effects because the number of trials was too small to estimate these variances and covariances with sufficient precision and concern that single random effects may not represent trial-level heterogeneity sufficiently well. Because baseline levels of depressive symptoms may well influence the symptom trajectory of internalizing, we regressed the slope on the intercept to control for baseline internalizing symptoms. A test of moderation of the intervention effect on slope trajectory by baseline level of depressive symptoms was

Fig. 2 Follow-up rates across time. Trial 18 is missing 49 participants at baseline



conducted by testing whether the regression coefficients of the latent slope on the latent intercept were different for intervention versus control.

Testing for Main Effects of Intervention on Depression Trajectories

In the second-order growth model, main effects of the intervention on the trajectory of adolescent depressive symptoms were tested by regressing the latent slope on the intervention condition; see equations in Brincks et al. (2016). This was coded so that a negative coefficient indicated more rapid reduction for intervention compared to control conditions. We tested the overall effect of intervention versus control using the estimated coefficient and a Wald test based on robust standard errors from this analysis. An effect size (ES) is calculated for these effects by dividing the difference in intervention versus control slopes with the pooled standard deviation of the slopes after adjustment for the intercept and baseline covariates. Negative ESs imply that symptoms improved more for intervention than for control.

Testing for Variation in Intervention Impact on Depression Trajectories

We conducted a number of tests to examine variation in impact. Moderation of the intervention effect on slope trajectory by baseline level of depressive symptoms was examined by testing whether the regression coefficients of the latent slope on the latent intercept differed by intervention condition; see equations in Brincks et al. (2016). All individual-level moderator effects mentioned earlier were tested similarly by Wald-type tests of the interaction between the intervention condition and the moderator of interest.

We were also interested in characteristics of the intervention that may moderate treatment effects, such as type of intervention (CBT, IPT, and PFS), target of the intervention (child, parent, conjoint), and whether the trial specifically targeted depression as an outcome. To conduct these analyses, we attempted two-level growth modeling (i.e., three-level mixed effects modeling involving time, person, and arm of trial) in Mplus (Muthén and Muthén 1998–2015). As these analyses did not converge due to the modest number of trial arms (24) fit with two correlated random effects, we instead estimated intervention effects using the second-order growth model with slope regressed on each trial (adjusting for intercept and individual-level covariates) and extracted these adjusted empirical Bayes estimates and their standard errors into a separate dataset with adjustments for trials that had two active intervention arms compared to the same control, as described in Brincks et al. (2016).

In addition to formal testing for moderation effects of individual-level and arm-level predictors, we examined

variation in impact three other ways. First, we conducted growth mixture models to assess whether impact varied by different patterns in growth of symptoms over time (Muthén et al. 2002). Second, we summarized the distribution of arm-level effects graphically in a smooth histogram by simulating the distribution of the contrasts using the empirical Bayes estimates and their standard errors. Third, we summarized variations in intervention versus control differences in impact using two-level regression analysis both as single moderator effects as well as effects adjusting for other potential moderators.

Mplus version 7.3 (Muthén and Muthén 1998–2015) was used to evaluate all structural equation models using comparative fit indices (CFI) and root mean square error of approximation (RMSEA) to assess overall model fit and chi-square and Wald tests to compare nested models. We use as guides to good model fitting the rules of Browne and Cudeck (RMSEA < 0.05; 1993) and Hu and Bentler, (CFI > 0.90; 1999). R (R Core Team 2015) was used to conduct orthogonal transforms and simulations.

Results

Balance across Intervention Condition

To compare baseline characteristics between the intervention and control groups, we performed Mantel–Haenszel tests of conditional independence for binary variables and random intercept multilevel model testing for continuous variables. As shown in Table 2, the numbers of subjects in these 19 trials were far from evenly divided by intervention ($N = 3098$) and control ($N = 2112$) groups, reflecting the fact that five trials had two active arms and one control and a few trials intentionally randomized more youth to intervention rather than control. The proportion of youth excluded from our analyses as a result of age (younger than 11 or older than 18) differed modestly across the intervention condition (5.5 % control participants excluded versus 6.9 % intervention participants excluded, $\chi^2 = 4.12, p = 0.04$). After removing these observations, there were a total of 5210 observations. No significant differences by intervention condition were found on the average number of times measured (4.06 times for the intervention group and 3.81 times in the control group, $p = 0.77$) nor on the average number of different symptom measures assessed (the intervention group averaged 1.70 of the eight different depressive symptom measures, while the average in the control group was 1.81, $p = 0.11$). Overall, there were no differences in baseline levels of gender, race/ethnicity, parent education level, or whether income information was missing. However, the intervention group reported a higher mean household income, US\$43,000, compared to US\$39,000 in the control group, an effect size difference of 0.13 ($p < 0.001$). Also, the intervention group was less likely to have a missing value on parent education level compared

Table 2 Baseline comparisons by intervention condition

	Intervention		Control		Comparison		
	Score	<i>df</i>	<i>p</i>	Score	<i>df</i>	<i>p</i>	
<i>N</i>	2868		2342				
Age	12.74	(1.51)	12.97	(1.55)	5.32 ^b	5208	<0.001
Gender	1434	(50.0 %)	1185	(50.6 %)	0.45 ^c		0.65
Race/ethnicity					3.24 ^d	4	0.52
Hispanic (all races)	923	(32.2 %)	780	(33.3 %)			
Black, non-Hispanic	31	(1.1 %)	24	(1.0 %)			
White, non-Hispanic	124	(4.3 %)	115	(4.9 %)			
Asian, non-Hispanic	0	(0.0 %)	2	(0.1 %)			
Other/multiple race	6	(0.2 %)	7	(0.3 %)			
Income (in US\$)	43	(33.00)	39	(31.00)	-4.10 ^b	3787	<0.001
Parent education					8.28 ^d	4	0.08
No high school	162	(5.6 %)	134	(5.7 %)			
Some high school	203	(7.1 %)	179	(7.6 %)			
High school/GED	536	(18.7 %)	365	(15.6 %)			
Some college	692	(24.1 %)	389	(16.6 %)			
College/higher degree	757	(26.4 %)	421	(18.0 %)			
Measures							
RBPC-ANX/WIT	5.33	(5.42)	4.97	(4.91)	-1.00 ^c		0.32
YSR-ANX/DEP	5.12	(4.57)	4.95	(4.32)	-0.84 ^c		0.4
YSR-WIT/DEP	3.96	(2.86)	3.93	(2.74)	-0.13 ^c		0.9
CDRS-R	28.61	(8.35)	28.62	(7.98)	0.03 ^c		0.98
CESD	18.47	(10.25)	18.77	(11.19)	0.43 ^c		0.67
CDI	8.95	(7.59)	9.35	(7.82)	1.19 ^c		0.23
CBCL-ANX/DEP	3.98	(3.65)	3.85	(3.50)	-0.53 ^c		0.59
CBCL-WIT/DEP	2.86	(2.50)	2.79	(2.53)	-0.39 ^c		0.69
Single-factor score ^a	0.002	(0.036)	0 (reference)				0.95

Values are reported as the mean (standard deviation) and *n* (%). Additionally, measures are listed with abbreviated titles as described in the text

^a Beta coefficient (standard error)

^b *t* scores

^c *z* scores

^d Chi-square

to the control (23.6 versus 29.7 %, Mantel–Haenszel $\chi^2 = 24.8$, $p < 0.001$). Lower scores on income and parent education are associated with higher rates of depressive symptoms, and therefore these two variables are included as covariates in all analyses.

Comparison of Baseline Level of Depression Symptoms

Our conclusions about intervention impact on the course of depression would be hampered if there were differences between intervention and control at baseline or if the underlying factor model of depressive symptom scores differed by intervention condition. We therefore began with a comparison of means on each of the observed baseline measures. Overall, there were no significant differences in the baseline levels of any of the eight depressive symptom measures rated by the

child (CDI, YSR-Anx/Dep, YSR-Wit/Dep, and CESD), parent (RBPC-Anx/Wit, CBCL-Anx/Dep, and CBCL-Wit/Dep), or clinician (CDRS; see Table 2).

Because of the high proportion of missing data on each of these individual variables, we conducted an overall test based on a single-factor model (see last row of Table 2). A model that enforced similar factor structures at baseline demonstrated adequate fit to the data ($\chi^2_{(288)} = 772.18$, $p < 0.001$; CFI = 0.88, RMSEA = 0.018). The model that allowed the latent variable’s loadings on the eight depressive symptom measures to differ by intervention condition did not improve the fit ($\chi^2_{(8)} = 9.15$, $p = 0.33$), further suggesting comparability. The factor model indicated that the eight proposed indicators of depression symptoms all loaded significantly onto a single latent construct. The standardized factor loadings at baseline

ranged from a low of 0.18 for RBPC anxiety/withdrawal to a high of 0.999 for CESD. The single clinician measure of CDRS-R had a standardized loading of $\lambda = 0.35$, which was much lower than CDI ($\lambda = 0.84$), YSR-ANX/DEP ($\lambda = 0.89$), and YSR-WIT/DEP ($\lambda = 0.89$). This indicated more measurement error for one parent report, moderate measurement error for the clinician rating and two parent ratings, and the lowest measurement error for the four self-reported measures. Based on this latent variable model, we computed the average level of depressive symptoms at baseline across all subjects. The baseline mean on the CESD scale was 13.42 (SD = 9.35), a few points below the clinical level of 16, indicating mild depression (Radloff 1977). This translates to approximately 40 % of our sample having symptom levels above a mild clinical level at baseline on the CESD scale.

To test for baseline differences in depressive symptoms, the latent factor was regressed on the intervention condition after adjusting for individual-level covariates of age, gender, race/ethnicity, family income, and parent education level. As expected in randomized trials, the baseline level of depression did not significantly differ by intervention condition ($b = 0.002$, $SE = 0.036$, $p = 0.95$), nor did this conclusion change when trials were entered as fixed factors.

Associations between Covariates and Baseline Depressive Symptoms

The baseline level of depressive symptoms was marginally significantly and positively related to age ($b = 0.037$, $SE = 0.019$, $p = 0.055$) and negatively related to being male ($b = -0.267$, $SE = 0.036$, $p < 0.001$), household income ($b = -0.036$, $SE = 0.008$, $p < 0.001$), and parent education level ($b = -0.050$, $SE = 0.021$, $p = 0.016$). After adjustment for trial, there was no significant difference in the baseline symptom

scores for Hispanics, African Americans, or white non-Hispanics ($\chi^2 = 0.367$ on 2 *df*, $p = 0.83$).

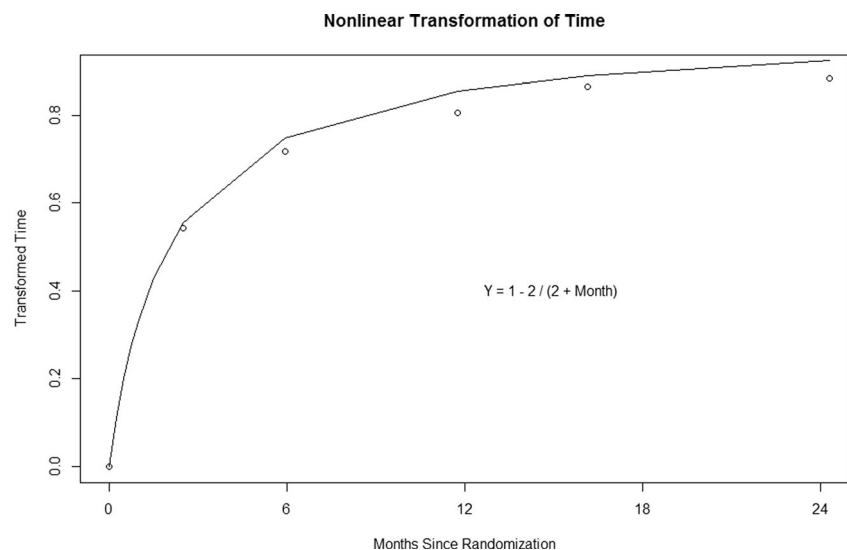
Second-Order Latent Growth Modeling

In this section, we first discuss the general pattern of growth trajectories we found across time and trial. In our second-level growth modeling analysis involving time points 0 through 6, leaving out time point 1 as discussed above, we allowed for potential nonlinear growth by fixing the first two loadings, then estimating the loadings for time points 3–6. The estimated loadings provide a general shape of a curve of the general pattern of symptom change over time, shown in Fig. 3, which shows the plot of time against the optimal transformed scale of time. Note that the curve very closely resembles an inverse transformation of time rather than being linear. Thus, depending on the sign of the slope coefficient from our analyses, the pattern of change would show an immediate rise in symptoms (for a positive slope) or an immediate lowering of symptoms (for a negative slope), followed by a slower rise or fall to an asymptote that is reached over 2 years. In our analyses below, regressing the latent slope on the intervention condition will result in a negative coefficient if those in the intervention group show a faster decrease in symptoms compared to controls. This corresponds in a mixed effects growth model to an interaction term between the intervention condition and this transformed timescale. These loadings were estimated in the main effects model, and the values were fixed to these estimates for all subsequent analyses (except growth mixture models) to provide direct comparability across analyses.

Intervention Effect on Change in Depressive Symptoms

The next analysis examined the overall effect of intervention on the change in depressive symptoms over the 24-month

Fig. 3 Functional transformation of time used in growth modeling



follow-up period. In this analysis, the depression factor loadings and intercepts were constrained to be equal across time points. Because we detected no significant differences at baseline and we wanted to make comparisons across intervention condition, we constrained the relationship between the latent intercept as well as all covariates to be equal for the intervention and control conditions. We also constrained the relationship between the latent trajectory and the latent baseline level (intercept) to be equal across the intervention condition. In these analyses, trial membership was treated as a fixed effect with 18 dummy-coded variables included as predictors of the intercept and slope, and fixed equal across the intervention condition. There was a significant mean difference in the trajectory of depressive symptoms ($b = -0.439$, $SE = 0.224$, Wald test = 3.852, $df = 1$, $p = 0.050$), with a larger reduction in depressive symptoms over time for those receiving intervention compared to control. Across all the trials, this overall effect size is 0.09 (95 % CI = 0.00–0.19).

Moderators of Intervention Effect

Our next series of analyses tested for moderation of intervention effects by (1) characteristics of the participants, including baseline measures of depressive symptoms, age, gender, and race/ethnicity; (2) family income and parent education; and (3) trial characteristics, including whether the intervention targeted depression directly.

Baseline Depression

To examine whether baseline depression moderated the intervention effect on the trajectory of depressive symptoms, we first tested whether the regression of the slope on intercept differed by intervention condition. Baseline depressive symptoms was a significant, negative predictor of change in depression symptoms across both intervention and control conditions (control: $b = -0.39$, $SE = 0.04$, $p < 0.001$; intervention: $b = -0.41$, $SE = 0.04$, $p < 0.001$), and this relationship between intercept and slope did not vary by intervention condition (Wald test = 0.42, $df = 1$, $p = 0.52$), indicating no moderation effect of baseline depression severity. These results suggest that higher symptom levels at baseline led to faster reductions in depressive symptoms for both intervention and control participants, a point we will elaborate in later analyses presented in this paper.

Age

We found no significant relationships between participant age and trajectory of depressive symptoms for both control youth ($b = 0.17$, $SE = 0.14$, $p = 0.24$) and intervention youth ($b = 0.10$, $SE = 0.14$, $p = 0.47$). These effects were not significantly

different across conditions (Wald test = 0.21, $df = 1$, $p = 0.65$), indicating no moderation effect by age.

Gender

Gender was not a significant predictor of trajectory of depressive symptoms among control youth ($b = -0.37$, $SE = 0.30$, $p = 0.23$). However, among intervention youth, males experienced significantly faster declines in symptoms compared to females ($b = -1.21$, $SE = 0.28$, $p < 0.001$). The relationship between gender and trajectory of depressive symptoms was significantly different across conditions; that is, males benefit more from intervention than do females (Wald test = 4.33, $df = 1$, $p = 0.037$).

Race/Ethnicity

We found no significant moderation of overall intervention impact compared to control by race or ethnicity across Hispanic, white non-Hispanic, African-American, or other minority (Wald test = 4.242, $df = 3$, $p = 0.24$).

Family Income

Income did not moderate the intervention impact ($b = 0.014$, $SE = 0.073$, $p = 0.85$).

Parent Education

Education also did not moderate the intervention impact ($b = 0.134$, $SE = 0.861$, $p = 0.407$).

Intervention Target

Among the 19 trials in this synthesis, nine directly targeted depression as a primary outcome, whereas the remainder were other-focused trials that targeted problem behavior or general mental health. Given this important distinction in trial focus, we examined intervention effects within these two subgroups of trials. We note first that there were important differences among the participants in these two types of trials. Participants in the depression-focused trials had much higher baseline levels of depressive symptoms compared to those in the problem behavior or general mental health trials. Specifically, those in the depression-focused trials started a full 1.59 standard deviation higher in symptoms than those in the problem behavior or general mental health trials. We also discovered baseline differences on measures of socioeconomic status across these two types of trials. The mean household income among participants in the depression-focused trials was US\$61,854 ($SD = \$30,019$) and only US\$30,882 ($SD = \$28,081$) among participants in the problem behavior or general mental health trials. Thus, on average, those in the

depression-focused trials lived in families with twice as high incomes. Similarly, 98 % of the parents in the depression-focused trials had at least a high school degree, while only 74 % of parents in the problem behavior or general mental health trials completed a high school degree. Of parents in the depression-focused trials, 82 % had attended at least some college, but only 46 % of the parents in the problem behavior or general mental health trials had attended college. Thus, similar to the baseline level of depressive symptoms and income, there is a large difference in the distribution of education by whether trials are depression-focused or problem behavior or general mental health ($\chi^2 = 841.49$, $df = 4$, $p < 0.001$).

After controlling for all baseline measures, we found that there was a significant and strong intervention effect among depression-focused trials. Those in the control condition in the nine depression-focused trials showed decreasing levels of symptoms overall, but those in the depression-focused intervention decreased more rapidly over the 2 years (second highest curve). The differences in the trajectories of the depression-focused trials were significant ($b = -1.093$, $CI = -1.763$ to -0.422 , $p = 0.001$) and of substantial magnitude ($ES = -0.239$). In the problem behavior or general mental health trials, the mean trajectory of depressive symptoms was not significantly different across conditions ($b = 0.034$, $CI = -0.515$ to 0.583 , $p = 0.903$); both intervention and control equally reduced symptoms over time.

Given the much higher baseline depressive symptoms for the depression-focused trials, we conducted a growth mixture analysis (Muthén et al. 2002) to examine whether depression-focused intervention conditions reduced symptoms more for those with high baseline depressive symptoms compared to the control. We found two distinct classes separated by their baseline symptom level. About 40 % of the sample was in the high class with average symptom levels 2–4 points below the clinically significant cutoff value of 20 on the CDI. The proportion in trials that focused on depression who fell in this high baseline group was 45 %, while those in the problem behavior- and mental health-focused interventions was slightly lower at 38 %, which was only marginally lower ($p = 0.05$). For these high-symptom youth, the depression-focused interventions reduced symptoms significantly ($b = -2.01$, $SE = 1.02$, $p = 0.050$). There was no intervention effect among those with high levels of symptoms at baseline who were in the problem behavior or general mental health interventions, nor was there any impact for either depression-focused or problem behavior or general mental health trials among those who started with low levels of depressive symptoms. We also replicated these growth mixture analysis findings by fitting separate growth models for those who had at least one baseline depression scores above the 60th percentile and for those who had no high baseline symptom scores. Figure 4 shows the changes across time in symptoms, calibrated against the CDI

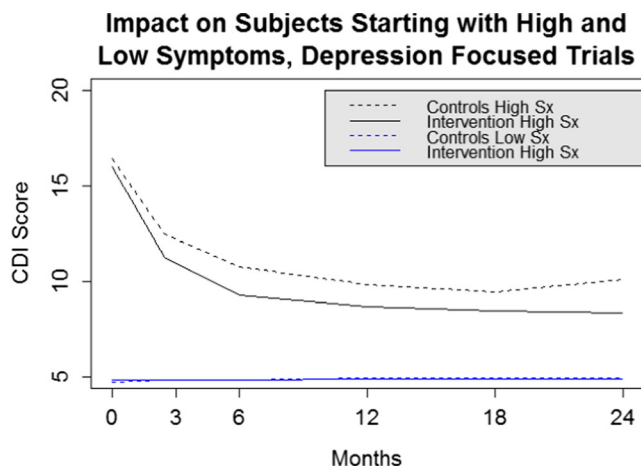


Fig. 4 Depressive symptoms across time among youth with high symptoms at baseline in the depression-focused trials and among youth with low symptoms at baseline

scale even though not all trials used this measure (see details in Brincks et al. 2016). The two uppermost curves indicate sustained impact of depression-focused prevention programs over control conditions for those who started with high symptoms. The bottommost curves in Fig. 4 show no impact for any of those who begin with low symptoms; there is also no elevation in these symptoms across time ($b = -0.105$, $SE = 0.15$, $p = 0.47$).

We looked further to see if there were any variations among impacts involving depression-focused interventions on youth with high levels of symptoms. In particular, we examined if there was a difference in impact among the depression-focused interventions for those that targeted the child directly or not (i.e., only targeted the parents). Among this highest-risk group at baseline, programs that specifically targeted the youth were much more effective in reducing symptoms than those targeting parents alone ($b = -3.525$, $SE = 0.888$, $p < 0.001$).

Moderation of Intervention Effects by Type of Intervention

Among the 19 trials, there were five trials that had two active intervention arms. Thus, a total of 24 active interventions were used in these trials. We were unable to create computationally accurate second-order growth models that accounted for all arm-level factors in one single analysis due to the large amount of missing data and relatively few arms. Instead, we calculated arm-level intervention versus control impact on the slope of depressive symptoms for each of the 24 arms using dummy variables for arm and adjusting for age, gender, race/ethnicity, parent education, family income, and baseline levels of depressive symptoms in each trial. The 24 arm-level interactions between each active intervention and control were then entered into a next stage analysis, using as input the empirical Bayes estimates, their standard errors, and, in the case of the five trials with two active

intervention arms, their correlations. A summary of the 24 effects is shown in a smoothed density plot in Fig. 5 that takes into account the uncertainty in each of these estimates. The effects on this figure are scaled in multiples of the overall standard deviation of the slope, i.e., effect sizes. Seventy percent of the mass is below zero, indicating an overall beneficial effect of these interventions compared to their respective controls. Nearly half (49.7 %) of the mass is below -0.5 standard deviation; one third of the mass shows a large beneficial effect size, more extreme than -1 standard deviation. On the other end of the distribution, 15 % of the mass is above the 0.5 standard deviation.

These empirical Bayes summary statistics were analyzed in a two-level analysis in order to examine arm-level effects both singly and in combination. We focused on six indices of intervention type using six dummy codes for whether the intervention included components that were delivered to the adolescent, delivered to the parent, or delivered conjointly, as well as whether the theoretical orientation of the intervention involved CBT, IPT, or PFS. These were not independent contrasts: for example, among the 18 arms where the child received intervention directly, 10 of these also included a parent-focused intervention. In addition, theoretical orientation was associated with participant type, with CBT interventions much more likely to include only the adolescent.

We first specified a model that included all six variables. None of the effects were significant, indicating that none of these intervention-level attributes independently accounted for differences in intervention effect. We then conducted exploratory analyses with each variable separately. Child-focused interventions showed a large reduction in symptoms over their controls compared to those that were not child-focused ($b = -1.418$, $SE = 0.502$, $p = 0.005$). No other individual moderator effects were significant.

Discussion

Our analyses demonstrated significant overall impact across the 19 trials of intervention versus control on depressive

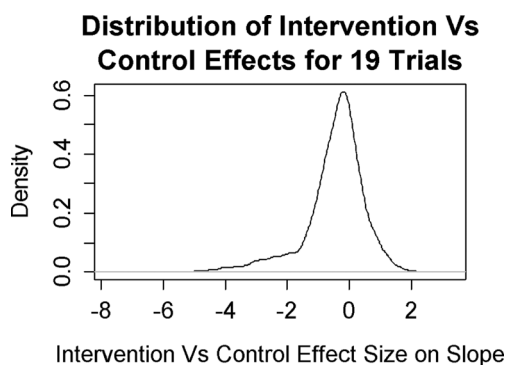


Fig. 5 Distribution of intervention versus control effect from 24 interventions in 19 trials

symptoms 2 years post-randomization. Collectively, these interventions significantly reduced depression symptoms over the first year, and reduction of symptoms was maintained up to 2 years after the intervention. This pattern of stable symptom reduction is exactly what one would want to see from indicated prevention programs working with adolescents having elevated symptoms of depression at baseline. In the intervention group, the reduction in symptoms was 10 % stronger than that of the control group relative to the standard deviation in the control group. As a whole, these interventions showed a preponderance of beneficial effects, with half of the preventive effect sizes above 0.5 standard deviation. This finding supports those of previous meta-analyses of interventions to prevent youth depression (Horowitz and Garber 2006; Merry et al. 2011; Stice et al. 2009). The results add to these existing findings by including problem behavior or general mental health trials besides depression-focused trials in the analyses.

There was evidence that these interventions varied in their impact and showed different effects on those with higher levels of symptoms. We found that interventions that specifically targeted depression, most of these having cognitive behavioral or interpersonal prevention components, had a beneficial effect ($ES = -0.24$), whereas those not specifically focusing on depression showed no overall effect. The impact of these depression-focused interventions over 2 years was limited to those who began with high levels of symptoms. In the absence of intervention, these youth remained near the clinically significant level, whereas those given a depression-focused intervention showed a reduction in their symptoms. We detected no change among youth whose depressive symptoms began at very low levels.

Our effect sizes are on the same order as those reported from meta-analyses. Horowitz and Garber (2006), whose meta-analysis of 30 trials that included a broader range of interventions than we selected here in terms of age of youth but a more limited variation in type of intervention, reported an overall ES of -0.16 at follow-up (typically 6 months), compared to our -0.09 across 2 years, and their selective and indicated ES s of -0.30 and -0.23 , respectively, were close to our ES of -0.24 for depression-focused trials among youth with high baseline depression scores. Stice et al. (2009), examining 47 trials, reported similar values, as did Merry et al. (2011) who reported a 12-month ES of -0.20 for targeted interventions and Sandler et al. (2014) who conducted an overview of existing meta-analyses, obtaining an ES of 0.19 for effects up to 1 year.

Our conclusion about the length of benefit of these prevention programs is stronger than that suggested by existing reviews (Horowitz and Garber 2006; Merry et al. 2011). Merry concluded that there were no effects by 24 months, but a modest effect at 36 months. Even though the meta-analyses are based on a larger numbers of trials than we had, their proportion of trials reporting data at specific times is small

and cannot take time into account the way we did with growth curve analyses, nor the added precision of using individual-level covariates.

In addition to reporting continuing impact through 2 years, our integrative data analysis was able to examine individual-level moderation by age, gender, or race/ethnicity, which was not available to these meta-analyses. We found consistent evidence that males benefitted more from intervention than did females. This is different from a previous review (Garber and Downs 2011) that noted a non-significant advantage for males and ecologic analyses using percent of females in the trial (Horowitz and Garber 2006; Stice et al. 2009), which suggest that trials with greater proportions of female participants show stronger intervention response. In our synthesis, there was no evidence of moderation by age, race/ethnicity, education level, or family income. We do not consider the question of moderation effects across these variables as closed, but this provides some evidence that youth of different ages and from different racial/ ethnic backgrounds seem to respond similarly to these preventive interventions. Because few studies measured parental psychopathology, we were not able to test whether there was a robust variation in impact by level of parental depression, as reported by Garber et al. (2009).

Because some of the problem behavior or general mental health interventions (i.e., those not specifically targeting depressive symptoms) have previously reported significant findings on reduction of symptoms of depression (Connell and Dishion 2008; Perrino et al. 2014; Sandler et al. 2003; Trudeau et al. 2012; Wolchik et al. 2013), we also included in our synthesis a set of these trials. The synthesis analyses suggest that, unlike the depression-focused interventions, these problem behavior or general mental health interventions did not show a significant reduction in overall effects on depressive symptoms compared to controls over the 2-year post-randomization period. Several issues should be noted in interpreting these findings. Overall, adolescents entered these problem behavior or general mental health intervention trials with substantially lower depressive symptoms as compared with those in the depression-focused interventions. In addition, both intervention and control adolescents in the problem behavior or general mental health interventions maintained low levels of depressive symptoms comparable to those of the depression-focused intervention group through the 2-year follow-up. This suggests that adolescents in the problem behavior or general mental health trials were, as a group, at much less risk of depression during the period of development studied here, and there was little room for these interventions to show improvement. Even for those who had elevated symptoms at baseline, exposure to interventions that did not specifically targeted depression showed a non-significant effect on the course of depressive symptoms. We note that our null findings through 2 years on depressive symptoms for problem behavior or general mental health preventive interventions do

not account for the potential for later impact, as others have found at 3 years (Connell and Dishion 2008) and 15 years (Wolchik et al. 2013). Follow-up periods longer than 2 years examined in the present analyses may be necessary to capture the potential cascading effects of problem behavior or general mental health interventions.

Most parenting/family programs in this set of studies were designed to prevent problem behavior or promote mental health, and indeed most of these studies have reported effects on these outcomes within the 2-year follow-up studied here. There is evidence that problem behaviors constitute a separate pathway to subsequent depression, with depression emerging only after these behaviors have become established and disrupted adolescent development (Wolchik et al. 2013). This reflects the “dual failure model” which suggests that, across time, early behavioral and conduct problems contribute to poor social and academic competence, which subsequently increases susceptibility to depressive or internalizing symptoms (Capaldi 1992; Capaldi and Stoolmiller 1999; Moilanen et al. 2010). Depression has been described as having complex and multifactorial causes (Garber 2006), for which different risk groups and pathways exist. Indeed, longitudinal mediation analyses of family-focused parenting programs have found support for both earlier effects on internalizing problems and effects on academic problems as mediating pathways leading term effects on depression symptoms (Sandler et al. 2015; Trudeau et al. 2012). Further research is needed to identify how intervention effects to change earlier developmental processes have a cascading effect to impact the long-term development of depression. As such, preventive interventions that target the most relevant risk and protective factors for distinct at-risk groups may be needed. An additional pathway for these “other”-focused interventions may be on maternal depression, the relief of which can have a salutary effect on their children’s depressive trajectories.

In addition, other evidence suggests that the samples in the depression-focused studies had very different risk profiles compared to those in the problem behavior or general mental health preventive trials. Most of the depression-focused programs selected adolescents who had elevated but subclinical depressive symptoms at baseline, and some trials had a substantial percentage of adolescents with prior history of major depressive episodes (Clarke et al. 2001; Garber et al. 2009).

It is also possible that the preventive impact of these interventions may occur for youth having baseline risk factors for depression other than elevated depressive symptoms. For instance, Perrino et al. (2014), in analyzing three Familias Unidas trials that varied by adolescent externalizing risk status, found evidence that adolescents in families having poor parent–adolescent communication at baseline demonstrated reductions in internalizing symptoms following this family/parenting intervention. There was no evidence for such effects in families with good communication. Again, this supports the

existence of diverse risk pathways contributing to adolescent depression. Indeed, poor parenting and family relationships can place youth at risk of depression (Restifo and Bogels 2009), and as such, a direct focus on improving parent–child communication provided by parent and family interventions may address different key risk processes for depression in some youth. It was not possible to examine family communication as a moderator across all 19 trials because parent–child communication was not measured in all these trials.

Methodologic Strengths and Limitations

These findings represent what we believe is the largest analysis of individual-level data on the prevention of depression. Analysis of individual-level data provides a much richer opportunity for synthesizing findings than does meta-analysis. In particular, IDA produces new analyses of individual-level data, unlike meta-analyses, which must rely on the varied analytic decisions of each research team that contributes to the published statistics. Finally, we had access to detailed information about these trials from those most familiar with these studies. This collaborative data synthesis study worked to involve principal investigators and research teams from the original individual trials to help guide and confirm the quality of the data, analyses, and interpretations (see Perrino et al. 2013 for further description of this collaborative data synthesis study).

We view one strength of this synthesis is the diverse set of interventions that are examined. It is feasible and appropriate to use IDA on a more homogeneous set of interventions than our set (e.g., one that was limited to depression-focused preventive interventions), but despite our wider trial inclusion, we were able to pull out the much stronger effect that the depression-focused trials had on those with high baseline symptoms.

This study has several limitations that need to be kept in mind. Our sample of trials does not include all possible trials of depression-focused and problem behavior or general mental health preventive interventions for adolescents. Nevertheless, our comparison with other published trials (supplemental material) and a comparison with effect sizes for trials identified in meta-analyses (Brincks et al. 2016; Sandler et al. 2014) did not reveal any obvious selection bias. Our set of trials is not fully representative. We did cover a diverse set of populations and interventions in the set of trials to improve power for testing moderation (Brown et al. 2013), and although we included a number of trials that focused on Hispanics, we did not include any that were focused only on African-American or other populations that differed from the dominant culture group. We have argued that the pool of trials focusing solely on minority populations, as well as the percentage of minorities in the trials themselves, still remain

lower than that needed to produce sufficient evidence of impact across these different populations (Perrino et al. 2015).

Another limitation is the lack of data on parental depression on many trials; parental depression has been shown to be a negative moderator of intervention outcome for CB treatment as well as prevention of depression (Brent et al. 1998, 2015; Garber et al. 2009; Lewinsohn et al. 1998). We were unable to examine whether parent depression at baseline moderated the effects across these preventive interventions because this variable was only measured in a few trials. In addition, trials vary substantially in the way the moderators themselves were measured. We were able to harmonize single items of family income and parent education despite the fact that these were categorized differently across studies, but found it difficult to harmonize multi-item constructs like parenting because the items across studies were conceptually very different. As a result, the set of trials that meet all requirements for conducting an IDA moderator analysis would still be smaller than the set of trials that have tested relevant main effects. This points to potential benefits of the routine use of common measures in randomized trials, as now being promoted through the PhenX Toolkit (<https://www.phenxtoolkit.org/>).

Our total dataset also had a complex pattern of missing data, given that trials employed different measures of depressive symptoms and collected follow-up data at different times. The quantitative methods we developed to analyze these data and deal with this missingness are novel, and conditions under which their assumptions hold need to be explored more thoroughly. We discuss these assumptions and potential limitations more fully in Brincks et al. (2016) and Siddique et al. (2016).

The vast majority of the trials did not assess diagnoses of major depression disorder, so we do not report the impact on diagnoses averted by these interventions. Whereas we found important evidence of moderation effects at the individual and trial levels, we have not fully investigated impact across all subgroups. Additional studies are underway to examine the impact of these interventions by exposure to socioeconomic adversity. This is important given the substantial differences in participants' socioeconomic and educational disadvantage across types of prevention trial, as noted in the current paper's results section.

Further, as few of these trials directly tested two active interventions against one another, there is minimal evidence of the comparative effectiveness of the different intervention components and parent and/or youth targets. Some of these comparisons were difficult to interpret because the trials varied dramatically on adolescent baseline depressive symptoms, income, and parent education. Also, this paper did not examine hypothesized mediators, intervention dosage, or outcomes beyond 2 years, given that only 5 of the 19 trials had longitudinal data beyond this time period. Finally, we note that all but two of the trials delivered at least part of their intervention in a

group format. None of the trials included the group clustering identity in their datasets, nor did they account for clustering when analyzing their own trials, so we were unable to account for this clustering in our analyses. Potentially, such clustering could increase standard errors and reduce the level of significance of our findings.

Implications

There is still a lack of knowledge about exactly how much improvement we would expect in preventing depressive disorders at a population level with different prevention strategies. Much of our epidemiologic knowledge about adolescent depression is focused on treatment rather than prevention, and we discuss this first. Epidemiologically, more than 9 % of the adolescent population 12–18 years old experienced a major depressive episode within the past year (Substance Abuse and Mental Health Services Administration & Center for Behavioral Health Statistics and Quality 2014). Current levels of treatment are inadequate. Among these youth who are experiencing a disorder, just one third of those receive healthcare to treat their depression (Avenevoli et al. 2015). If we could guarantee that all such youth were provided the high-quality treatment, antidepressants and/or CBT or IBT, we could expect about two thirds of these individuals to have a substantial reduction in depressive symptoms (Brent et al. 1997; Gibbons et al. 2012a, b), compared to about a third of these experiencing a reduction in symptoms as a consequence of the natural episodicity of depression. Thus, under our current treatment system, we can expect well below 20 % of the more than 2 million 12- to 18-year-olds experiencing a major depressive episode to be provided an intervention that will ultimately improve their outcome. That leaves at least 1.6 million depressed youth without an effective treatment. Given the vast level of untreated adolescents with depression, prevention strategies can play a major role in addressing this unmet mental health need. Our analysis has shown that prevention is indeed effective in reducing depressive symptoms prior to the onset of a full depressive episode for those with subclinical symptoms, and while our trials did not afford the opportunity to assess averted depressive disorders, it is likely that they too were affected. Indeed, in this era of major health policy change to expand mental health coverage for treatment and prevention, the prevention programs that we examined here may provide a fundamental first line of defense against a disorder that, if untreated, often has lifetime consequences on mental health (Thapar et al. 2012; Weisz et al. 2005), physical health (Prince et al. 2007), and disability (Andrews et al. 2000; Ferrari et al. 2013). A growing area of both research and service involves how to integrate effective behavioral health into primary care (Asarnow et al. 2005; Butler et al. 2008; Jaycox et al. 2006; Wellset al. 2001), and we recommend that these efforts be expanded.

Attempts to prevent depression have ranged from universal interventions to highly targeted risk populations. Some but not all of these have been successful. Our analyses show that among those who have elevated subclinical symptoms, the effects of prevention programs, particularly those directly targeting the child, have a strong and lasting impact. We believe the evidence of their impact is sufficiently strong to recommend their expanded use in diverse settings that until now have focused mostly on treatment for depression (e.g., primary care, children's mental health, foster care, juvenile justice, and schools), if at all. The preventive interventions described here provide a major opportunity to improve the nation's mental health for the following reasons. First, the problem of youth depression is exceptionally high. Close to 10 % of youth experience a major depressive disorder (MDD) in the USA each year (Substance Abuse and Mental Health Services Administration & Center for Behavioral Health Statistics and Quality 2014), with 15 % having a diagnosis of MDD before age 25 and an additional 10 % experiencing a minor depressive disorder (Kessler and Walters 1998). Second, symptoms of depression usually arise two to three or more years before a diagnosis of MDD occurs, and generally around age 12 (Institute of Medicine & National Research Council 2009), thus providing an excellent opportunity to deliver effective prevention programs before first episode onset. Third, depression is an episodic disorder, with nearly three fourths reporting recurrence during childhood or young adulthood, with very few episodes lasting over 2 years (Kessler and Walters 1998). This episodicity provides multiple opportunities to offer preventive interventions, which have been shown to have long-term effects for those who have experienced prior episodes. Finally, the major expansion of mental health services through the Mental Health Parity and Addictions Equality Act and the Patient Protection and Affordable Care Act has now provided expanded mental health services for an estimated 62 million Americans (Frank et al. 2014), a substantial portion of whom are adolescents. The use of pediatrics and family medicine to deliver such programs seems very promising, although research is needed to test how these prevention programs can be most effectively delivered and financed within those settings. In particular, the current recommendations from the US Preventive Services Task Force to provide screenings and treatment for depression for those between 12 and 18 years of age could usefully and readily be expanded to include prevention-focused interventions for subclinical levels of depression.

While some parenting interventions have shown longer-term impact on depression beyond 2 years, overall, we did not observe an impact on depression, and thus their primary use is to prevent substance use, conduct disorder, violence, and sexual risk behavior (Leslie 2016). Because the parenting interventions we included in this paper were primarily designed to prevent externalizing problems, substance use, or

to promote adjustment to stressful transitions, it may be that they did not focus on those aspects of parenting or family life that most impact the development of depression. Yap and Jorm (2015), in a meta-analytic study, reported specific aspects of family functioning that are most strongly associated with depression and internalizing problems, including high interparental conflict, parental over-involvement, abusive parenting, and low parental warmth. Parent-focused components in prevention programs may be quite appropriate in specific situations, particularly when parents themselves are undergoing a depressive episode (Brent et al. 2015; Garber et al. 2009) and when low baseline levels of parent–child communication can be improved (Perrino et al. 2014). Family-focused intervention trials are underway to test whether changing these specific aspects of family life leads to prevention of youth depressive symptoms (Compas et al. 2009, 2011).

In addition, the current data found no diminishment of program effects for ethnic minorities compared to majority youth. This is important because ethnic and racial minority youth experience greater socioeconomic and educational disadvantage, important risk factors for the development of depression (Reiss 2013; Yoshikawa et al. 2012). Implementing preventive interventions in minority communities should be considered, particularly where there is limited access to, or use of, primary and mental healthcare and where mental disorder diagnosis is highly stigmatizing (Cummings and Druss 2011).

In conclusion, this paper demonstrates that prevention programs have a robust and persistent beneficial effect on adolescents' depressive symptoms, lasting up to 2 years. Youth who start with higher depressive symptoms (approaching but below clinical levels) and are given CBT or IPT interventions directly provide the best opportunity for long-term benefit.

Acknowledgments We gratefully acknowledge the National Institute of Mental Health Collaborative Synthesis for Adolescent Depression Trials Study Team, comprising our many colleagues who generously provided their data to be used in this study, obtained access to key datasets, reviewed coding decisions, provided substantive or methodologic recommendations, and reviewed this manuscript. We also thank NIMH for their support through grant number R01MH040859 (Collaborative Synthesis for Adolescent Depression Trials, Brown PI), and the following grants: Garber, Brent, Beardslee, Clarke et al.—NIMH MH064735, MH064503, MH064717; Gillham et al.—NIMH MH52270; Garber et al.—William T. Grant Foundation 961730; Dishion et al.—NIDA DA07031 and DA13773; Szapocznik et al.—NIMH MH61143; Pantin et al.—NIDA DA017462; Prado et al.—NIDA DA025894; Prado et al.—CDC U01PS000671; Stormshak et al.—NIDA DA018374; Sandler et al.—NIMH MH49155; Wolchik et al.—NIMH MH068685; Young et al.—NARSAD; Spoth et al.—NIDA DA 007029; Clarke et al.—NIMH MH 48118; Young et al.—NIMH MH071320; Beardslee et al.—NIMH MH48696; VanVoorhees et al.—NIMH MH072918; and Gonzales et al.—NIMH MH64707. The content of this paper is solely the responsibility of the authors and does

not necessarily represent the official views of the funding agencies or that of our collaborators who provided access to their data.

Compliance with Ethical Standards

Funding Funding for this project was by NIMH R01MH040859 (Collaborative Synthesis for Adolescent Depression Trials, Brown PI) and the following grants for the specific trials: Garber, Brent, Beardslee, Clarke et al.—NIMH MH064735, MH064503, MH064717; Gillham et al.—NIMH MH52270; Garber et al.—William T. Grant Foundation 961730; Dishion et al.—NIDA DA07031 and DA13773; Szapocznik et al.—NIMH MH61143; Pantin et al.—NIDA DA017462; Prado et al.—NIDA DA025894; Prado et al.—CDC U01PS000671; Stormshak et al.—NIDA DA018374; Sandler et al.—NIMH MH49155; Wolchik et al.—NIMH MH068685; Young et al.—NARSAD; Spoth et al.—NIDA DA 007029; Clarke et al.—NIMH MH 48118; Young et al.—NIMH MH071320; Beardslee et al.—NIMH MH48696; VanVoorhees et al.—NIMH MH072918; and Gonzales et al.—NIMH MH64707.

Conflict of Interest CHB was supported as a consultant on one of these projects (New Beginnings) and received funding support on the Familias Unidas trials, as did Huang, Young, Sandler, Pantin, and Beardslee directed their own trials that are included in this synthesis. Brincks, Perrino, and Cruden were colleagues of Pantin. All the trial directors were consulted on codings for their trials and were offered an opportunity to provide comments before this paper was submitted.

Ethics Approval This article does not contain any studies with animals performed by any of the authors. This project involving sharing of data was reviewed by two IRBs. Use of deidentified data in this synthesis project was approved by the University of Miami and Northwestern University IRBs, and all institutions signed data use agreements with Northwestern.

Informed Consent All trials included in this synthesis were approved by their respective institutional review boards. Informed consent was obtained from all individual participants included in these respective studies.

References

- Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and 1991 profile*. Burlington: University of Vermont Department of Psychiatry.
- Achenbach, T. M., & Rescorla, L. (2001). *The manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.
- Andrews, G., Sanderson, K., Corry, J., & Lapsley, H. M. (2000). Using epidemiological data to model efficiency in reducing the burden of depression. *The Journal of Mental Health Policy and Economics*, 3, 175–186.
- Asarnow, J. R., Jaycox, L. H., Duan, N., LaBorde, A. P., Rea, M. M., Murray, P., et al. (2005). Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics: A randomized controlled trial. *JAMA*, 293, 311–319. doi:10.1001/jama.293.3.311.
- Avenevoli, S., Swendsen, J., He, J.-P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the National Comorbidity Survey—adolescent supplement: Prevalence, correlates, and treatment.

- Journal of the American Academy of Child and Adolescent Psychiatry*, 54, 37–44 e32. doi:10.1016/j.jaac.2014.10.010.
- Beardslee, W. R., Gladstone, T. R. G., Wright, E. J., Cooper, A. B. (2003). A family-based approach to the prevention of depressive symptoms in children at risk: evidence of parental and child change. *Pediatrics*, 112(2), e119–131. doi:10.1542/peds.112.2.e119.
- Bertha, E. A., & Balázs, J. (2013). Subthreshold depression in adolescence: A systematic review. *European Child and Adolescent Psychiatry*, 22, 589–603. doi:10.1007/s00787-013-0411-0.
- Brent, D. A., Holder, D., Kolko, D., Birmaher, B., Baugher, M., Roth, C., et al. (1997). A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. *Archives of General Psychiatry*, 54, 877–885.
- Brent, D. A., Kolko, D. J., Birmaher, B., Baugher, M., Bridge, J., Roth, C., & Holder, D. (1998). Predictors of treatment efficacy in a clinical trial of three psychosocial treatments for adolescent depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 906–914. doi:10.1097/00004583-199809000-00010.
- Brent, D. A., Brunwasser, S. M., Hollon, S. D., Weersing, V. R., Clarke, G. N., Dickerson, J. F., et al. (2015). Effect of a cognitive-behavioral prevention program on depression 6 years after implementation among at-risk adolescents: A randomized clinical trial. *JAMA Psychiatry*, 72, 1110–1118. doi:10.1001/jamapsychiatry.2015.1559.
- Brincks, A., Huang, S., Brown, C. H., Howe, G. W., Pantin, H., & Perrino, T. (2016). Multilevel growth models for combining and synthesizing individual participant data from multiple prevention trials. *Prevention Science* (in press).
- Brown, C. H., Ten Have, T. R., Jo, B., Dagne, G., Wyman, P. A., Muthen, B., & Gibbons, R. D. (2009). Adaptive designs for randomized trials in public health. *Annual Review of Public Health*, 30, 1–25. doi:10.1146/annurev.publhealth.031308.100223.
- Brown, C. H., Sloboda, Z., Faggiano, F., Teasdale, B., Keller, F., Burhart, G.; Prevention Science and Methodology Group. (2013). Methods for synthesizing findings on moderation effects across multiple randomized trials. *Prevention Science*, 14, 144–156. doi:10.1007/s11121-011-0207-8.
- Butler, M., Kane, R. L., McAlpine, D., Kathol, R. G., Fu, S. S., Hagedorn, H., & Wilt, T. J. (2008). *Integration of mental health/substance abuse and primary care. Evidence Reports/Technology Assessments, No. 173. AHRQ Publication No. 09-E003*. Rockville, MD: Agency for Healthcare Research and Quality.
- Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up at grade 8. *Development and Psychopathology*, 4, 125–144.
- Capaldi, D. M., & Stoolmiller, M. (1999). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: III. Prediction to young-adult adjustment. *Development and Psychopathology*, 11, 59–84.
- Cardemil, E. V., Reivich, K. J., Beevers, C. G., Seligman, M. E. P., & James, J. (2007). The prevention of depressive symptoms in low-income, minority children: Two-year follow-up. *Behaviour Research and Therapy*, 45, 313–327. doi:10.1016/j.brat.2006.03.010.
- Clarke, G. N., Hawkins, W., Murphy, M., & Sheeber, L. (1993). School-based primary prevention of depressive symptomatology in adolescents: findings from two studies. *Journal of Adolescent Research*, 8, 183–204.
- Clarke, G. N., Hornbrook, M., Lynch, F., Polen, M., Gale, J., Beardslee, W., et al. (2001). A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Archives of General Psychiatry*, 58, 1127–1134.
- Compas, B. E., Forehand, R., Keller, G., Champion, J. E., Rakow, A., Reeslund, K. L., et al. (2009). Randomized controlled trial of a family cognitive-behavioral preventive intervention for children of depressed parents. *Journal of Consulting and Clinical Psychology*, 77, 1007–1020. doi:10.1037/a0016930.
- Compas, B. E., Forehand, R., Thigpen, J. C., Keller, G., Hardcastle, E. J., Cole, D. A., et al. (2011). Family group cognitive-behavioral preventive intervention for families of depressed parents: 18- and 24-month outcomes. *Journal of Consulting and Clinical Psychology*, 79, 488–499. doi:10.1037/a0024254.
- Connell, A. M., & Dishion, T. J. (2008). Reducing depression among at-risk early adolescents: Three-year effects of a family-centered intervention embedded in schools. *Journal of Family Psychology*, 22, 574–585. doi:10.1037/0893-3200.22.3.574.
- Cuijpers, P., van Straten, A., Smit, F., Mihalopoulos, C., & Beekman, A. (2008). Preventing the onset of depressive disorders: A meta-analytic review of psychological interventions. *American Journal of Psychiatry*, 165, 1272–1280. doi:10.1176/appi.ajp.2008.07091422.
- Cummings, J. R., & Druss, B. G. (2011). Racial/ethnic differences in mental health service use among adolescents with major depression. *Journal of the American Academy of Child and Adolescent Psychiatry*, 50, 160–170. doi:10.1016/j.jaac.2010.11.004.
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14, 81–100. doi:10.1037/a0015914.
- Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., et al. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research*, 49, 214–231. doi:10.1080/00273171.2014.889594
- Dagne, G., Brown, C. H., Howe, G., Kellam, S., & Liu, L. (2016). Testing moderation in network meta-analysis with individual participant data. *Statistics in Medicine*, 35, 2485–2502. doi:10.1002/sim.6883.
- Dishion, T. J., Kavanagh, K., Schneiger, A., Nelson, S., Kaufman, N. K. (2002). Preventing early adolescent substance use: a family-centered strategy for the public middle school. *Prevention Science*, 3(3), 191–201.
- Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., et al. (2013). Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010. *PLoS Medicine*, 10, e1001547. doi:10.1371/journal.pmed.1001547.
- Frank, R. G., Beronio, K., & Glied, S. A. (2014). Behavioral health parity and the Affordable Care Act. *Journal of Social Work in Disability & Rehabilitation*, 13, 31–43. doi:10.1080/1536710X.2013.870512.
- Garber, J. (2006). Depression in children and adolescents: Linking risk research and prevention. *American Journal of Preventive Medicine*, 31, S104–S125. doi:10.1016/j.amepre.2006.07.007.
- Garber, J., & Downs, L. (2011). Prevention of depression in youth: Sex differences in effects. In T. J. Strauman, P. R. Costanzo, & J. Garber (Eds.), *Depression in adolescent girls: Science and prevention* (pp. 193–234). New York: Guilford.
- Garber, J., Clarke, G. N., Weersing, V. R., Beardslee, W. R., Brent, D. A., Gladstone, T. R. G., et al. (2009). Prevention of depression in at-risk adolescents: A randomized controlled trial. *JAMA: Journal of the American Medical Association*, 301, 2215–2224. doi:10.1001/jama.2009.788.
- Gibbons, R. D., Brown, C. H., Hur, K., Davis, J. M., & Mann, J. J. (2012a). Suicidal thoughts and behavior with antidepressant treatment: Reanalysis of the randomized placebo-controlled studies of fluoxetine and venlafaxine. *Archives of General Psychiatry*, 69, 580–587. doi:10.1001/archgenpsychiatry.2011.2048.
- Gibbons, R. D., Hur, K., Brown, C. H., Davis, J. M., & Mann, J. J. (2012b). Benefits from antidepressants: Synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Archives of General Psychiatry*, 69, 572–579. doi:10.1001/archgenpsychiatry.2011.2044.
- Gillham, J. E., Reivich, K. J., Brunwasser, S. M., Freres, D. R., Chajon, N. D., Kash-MacDonald, V., et al. (2012). Evaluation of a group cognitive-behavioral depression prevention program for young adolescents: A randomized effectiveness trial. *Journal of Clinical*

- Child & Adolescent Psychology*, 41, 621–639. doi:10.1080/15374416.2012.706517.
- Gillham, J. E., Reivich, K. J., Freres, D. R., Chaplin, T. M., Shatte, A. J., Samuels, B., et al. (2007). School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn Resiliency Program. *Journal of Consulting & Clinical Psychology*, 75(1), 9–19. doi:10.1037/0022-006X.75.1.9.
- Gonzales, N. A., Dumka, L. E., Millsap, R. E., Gottschall, A., McClain, D. B., Wong, J. J., et al. (2012). Randomized trial of a broad preventive intervention for Mexican American adolescents. *Journal of Consulting and Clinical Psychology*, 80, 1–16. doi:10.1037/a0026063.
- Greenbaum, P. E., Wang, W., Henderson, C. E., Kan, L., Hall, K., Dakof, G. A., & Liddle, H. A. (2015). Gender and ethnicity as moderators: Integrative data analysis of multidimensional family therapy randomized clinical trials. *Journal of Family Psychology*, 29, 919–930. doi:10.1037/fam0000127.
- Grieco, E. M., & Cassidy, R. C. (2001). Overview of race and Hispanic origin: Census 2000 brief. Retrieved from <http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf>. Accessed 16 Dec 2016
- Higgins, J., Whitehead, A., Turner, R. M., Omar, R. Z., & Thompson, S. G. (2001). Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine*, 20, 2219–2241. doi:10.1002/sim.918.
- Hitlin, S., Scott Brown, J., & Elder, G. H. (2006). Racial self-categorization in adolescence: Multiracial development and social pathways. *Child Development*, 77, 1298–1308. doi:10.1111/j.1467-8624.2006.00935.x.
- Horowitz, J. L., & Garber, J. (2006). The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 74, 401–415. doi:10.1037/0022-006X.74.3.401.
- Horowitz, J. L., Garber, J., Ciesla, J. A., Young, J. F., & Mufson, L. (2007). Prevention of depressive symptoms in adolescents: A randomized trial of cognitive-behavioral and interpersonal prevention programs. *Journal of Consulting and Clinical Psychology*, 75, 693–706. doi:10.1037/0022-006X.75.5.693.
- Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*, 9, 61–89. doi:10.1146/annurev-clinpsy-050212-185522.
- Institute of Medicine, & National Research Council. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington, DC: The National Academies Press.
- Jaycox, L. H., Asarnow, J. R., Sherbourne, C. D., Rea, M. M., LaBorde, A. P., & Wells, K. B. (2006). Adolescent primary care patients' preferences for depression treatment. *Administration and Policy in Mental Health and Mental Health Services Research*, 33, 198–207. doi:10.1007/s10488-006-0033-7.
- Kessler, R. C., & Walters, E. E. (1998). Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the National Comorbidity Survey. *Depression and Anxiety*, 7, 3–14. doi:10.1002/(SICI)1520-6394(1998)7:1<3::AID-DA2>3.0.CO;2-F.
- Kovacs, M., & Beck, A. T. (1977). An empirical-clinical approach toward a definition of childhood depression. In: J. G. Schulterbrandt and A. Raskin (Eds.), *Depression in childhood: Diagnosis, treatment, and conceptual models* (pp. 1–25). New York: Raven.
- Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59, 877–883.
- Leslie, L. K., Mehus, C. J., Hawkins, J. D., Boat, T., McCabe, M.A., & Barkins, S. (2016). Primary health care: Potential home for family-focused preventive interventions. *American Journal of Preventive Medicine*, 51, S106–S118.
- Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1998). Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. *Clinical Psychology Review*, 18, 765–794.
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science*, 587, 69–81.
- Marchand, E., Ng, J., Rohde, P., & Stice, E. (2010). Effects of an indicated cognitive-behavioral depression prevention program are similar for Asian American, Latino, and European American adolescents. *Behaviour Research and Therapy*, 48, 821–825. doi:10.1016/j.brat.2010.05.005.
- McClain, D. B., Wolchik, S. A., Winslow, E., Tein, J.-Y., Sandler, I. N., & Millsap, R. E. (2010). Developmental cascade effects of the New Beginnings Program on adolescent adaptation outcomes. *Development and Psychopathology*, 22, 771–784. doi:10.1017/S0954579410000453.
- Substance Abuse and Mental Health Services Administration & Center for Behavioral Health Statistics and Quality. (2014). *The NSDUHREPORT: Substance use and mental health estimates from the 2013 National Survey on Drug Use and Health: Overview of findings*. Retrieved from <http://www.samhsa.gov/data/sites/default/files/NSDUH-SR200-RecoveryMonth-2014/NSDUH-SR200-RecoveryMonth-2014.htm>.
- Merry, S. N., Hetrick, S. E., Cox, G. R., Brudevold-Iversen, T., Bir, J. J., & McDowell, H. (2011). Psychological and educational interventions for preventing depression in children and adolescents. *Cochrane Database of Systematic Reviews*, (12), CD003380. doi:10.1002/14651858.CD003380.pub3.
- Moilanen, K. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and Psychopathology*, 22, 635–653. doi:10.1017/S0954579410000337.
- Munoz, R. F., Beardslee, W. R., & Leykin, Y. (2012). Major depression can be prevented. *American Psychologist*, 67, 285–295. doi:10.1037/a0027666.
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide (seventh edition)*. Los Angeles: Muthén & Muthén.
- Muthén, B. O., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, 3, 459–475. doi:10.1093/biostatistics/3.4.459.
- National Institutes of Health. (2001). *NIH policy on reporting race and ethnicity data: Subjects in clinical research*. Retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-01-053.html>.
- Pantin, H., Prado, G., Lopez, B., Huang, S., Tapia, M. I., Schwartz, S. J., et al. (2009). A randomized controlled trial of Familias Unidas for Hispanic adolescents with behavior problems. *Psychosomatic medicine*, 71(9), 987–995. doi:10.1097/PSY.0b013e3181bb2913.
- Perrino, T., Howe, G., Sperling, A., Beardslee, W., Sandler, I., Shern, D., et al. (2013). Advancing science through collaborative data sharing and synthesis. *Perspectives on Psychological Science*, 8(4), 433–444. doi:10.1177/1745691613491579.
- Perrino, T., Pantin, H., Prado, G., Huang, S., Brincks, A., Howe, G., et al. (2014). Preventing internalizing symptoms among Hispanic adolescents: A synthesis across Familias Unidas trials. *Prevention Science*, 15, 917–928. doi:10.1007/s11121-013-0448-9.
- Perrino, T., Beardslee, W., Bernal, G., Brincks, A., Cruden, G., Howe, G., et al. (2015). Toward scientific equity for the prevention of depression and depressive symptoms in vulnerable youth. *Prevention Science*, 16, 642–651. doi:10.1007/s11121-014-0518-7.
- Perrino, T., Brincks, A., Howe, G., Brown, C. H., Prado, G., & Pantin, H. (2016). Reducing internalizing symptoms among high-risk, Hispanic adolescents: Mediators of a preventive family intervention. *Prevention Science*, 17, 595–605.
- Petkova, E., Tarpey, T., Huang, L., & Deng, L. (2013). Interpreting meta-regression: Application to recent controversies in antidepressants'

- efficacy. *Statistics in Medicine*, 32, 2875–2892. doi:10.1002/sim.5766.
- Poznanski, E. O., & Mokros, H. B. (1996). *Children's Depression Rating Scale, Revised (CDRS-R)*. Los Angeles: Western Psychological Services Los Angeles.
- Prado, G., & Pantin, H. (2011). Reducing substance use and HIV health disparities among Hispanic youth in the U.S.A.: The Familias Unidas program of research. *Intervencion Psicosocial*, 20, 63–73. doi:10.5093/in2011v20n1a6.
- Prado, G., Pantin, H., Briones, E., Schwartz, S. J., Feaster, D., Huang, S., et al. (2007). A randomized controlled trial of a parent-centered intervention in preventing substance use and HIV risk behaviors in Hispanic adolescents. *Journal of Consulting and Clinical Psychology*, 75(6), 914–926. doi:10.1037/0022-006X.75.6.914.
- Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *The Lancet*, 370, 859–877. doi:10.1016/S0140-6736(07)61238-0.
- Quay, H. C. (1983). A dimensional approach to behavior disorder: The Revised Behavior Problem Checklist. *School Psychology Review*, 12, 244–249.
- R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401.
- Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science and Medicine*, 90, 24–31. doi:10.1016/j.socscimed.2013.04.026.
- Restifo, K., & Bogels, S. (2009). Family processes in the development of youth depression: Translating the evidence to treatment. *Clinical Psychology Review*, 29, 294–316. doi:10.1016/j.cpr.2009.02.005.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Sander, J. B., & McCarty, C. A. (2005). Youth depression in the family context: Familial risk factors and models of treatment. *Clinical Child and Family Psychology Review*, 8, 203–219. doi:10.1007/s10567-005-6666-3.
- Sandler, I. N., Ayers, T. S., Wolchik, S. A., Tein, J.-Y., Kwok, O.-M., Haine, R. A., et al. (2003). The Family Bereavement Program: Efficacy evaluation of a theory-based prevention program for parentally bereaved children and adolescents. *Journal of Consulting and Clinical Psychology*, 71, 587–600. doi:10.1037/0022-006X.71.3.587.
- Sandler, I. N., Wolchik, S. A., Cruden, G., Mahrer, N. E., Ahn, S., Brincks, A., & Brown, C. H. (2014). Overview of meta-analyses of the prevention of mental health, substance use, and conduct problems. *Annual Review of Clinical Psychology*, 10, 243–273. doi:10.1146/annurev-clinpsy-050212-185524.
- Sandler, I. N., Ingram, A., Wolchik, S., Tein, J. Y., & Winslow, E. (2015). Long-term effects of parenting-focused preventive interventions to promote resilience of children and adolescents. *Child Development Perspectives*, 9, 164–171.
- Siddique, J., de Chavez, P. J., Howe, G., Cruden, G., & Brown, C. H. (2016). Limitations in using multiple imputation to harmonize individual participant data for meta-analysis. *Prevention Science* (in press).
- Spoth, R. L., Redmond, C., Shin, C. (2001). Randomized trial of brief family interventions for general populations: adolescent substance use outcomes 4 years following baseline. *Journal of Consulting & Clinical Psychology*, 69(4), 627–642.
- Stewart, L., & Clarke, M. (1995). Practical methodology of meta-analysis using updated individual level data. *Statistics in Medicine*, 14, 2057–2079.
- Stice, E., Shaw, H., Bohon, C., Marti, C. N., & Rohde, P. (2009). A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. *Journal of Consulting and Clinical Psychology*, 77, 486–503. doi:10.1037/a0015168.
- Stormshak, E. A., Connell, A. M., Véronneau, M. H., Myers, M. W., Dishion, T. J., Kavanagh, K., et al. (2011). An ecological approach to promoting early adolescent mental health and social adaptation: Family-centered intervention in public middle schools. *Child Development*, 82(1), 209–225.
- Thapar, A., Collishaw, S., Pine, D. S., & Thapar, A. K. (2012). Depression in adolescence. *The Lancet*, 379, 1056–1067. doi:10.1016/S0140-6736(11)60871-4.
- Trudeau, L., Spoth, R., Randall, G. K., Mason, W. A., & Shin, C. (2012). Internalizing symptoms: Effects of a preventive intervention on developmental pathways from early adolescence to young adulthood. *Journal of Youth and Adolescence*, 41, 788–801. doi:10.1007/s10964-011-9735-6.
- Van Voorhees, B. W., Fogel, J., Reinecke, M. A., Gladstone, T., Stuart, S., Gollan, J., et al. (2009). Randomized clinical trial of an Internet-based depression prevention program for adolescents (Project CATCH-IT) in primary care: 12-week outcomes. *Journal of Developmental and Behavioral Pediatrics*, 30(1), 23–37. doi:10.1097/DBP.0b013e3181966c2a.
- Weisz, J. R., Sandler, I. N., Durlak, J. A., & Anton, B. S. (2005). Promoting and protecting youth mental health through evidence-based prevention and treatment. *American Psychologist*, 60, 628–648. doi:10.1037/0003-066X.60.6.628.
- Wells, K. B., Kataoka, S. H., & Asarnow, J. R. (2001). Affective disorders in children and adolescents: Addressing unmet need in primary care settings. *Biological Psychiatry*, 49, 1111–1120.
- Wesselhoeft, R., Sørensen, M. J., Heiervang, E. R., & Bilberg, N. (2013). Subthreshold depression in children and adolescents—A systematic review. *Journal of Affective Disorders*, 151, 7–22. doi:10.1016/j.jad.2013.06.010.
- Wolchik, S. A., West, S. G., Sandler, I. N., Tein, J.-Y., Coatsworth, D., Lengua, L., et al. (2000). An experimental evaluation of theory-based mother and mother-child programs for children of divorce. *Journal of Consulting and Clinical Psychology*, 68, 843–856. doi:10.1037/0022-006X.68.5.843.
- Wolchik, S. A., Sandler, I. N., Tein, J.-Y., Mahrer, N. E., Millsap, R. E., Winslow, E., et al. (2013). Fifteen-year follow-up of a randomized trial of a preventive intervention for divorced families: Effects on mental health and substance use outcomes in young adulthood. *Journal of Consulting and Clinical Psychology*, 81, 660–673. doi:10.1037/a0033235.
- Yap, M. B. H., & Jorm, A. F. (2015). Parental factors associated with childhood anxiety, depression, and internalizing problems: A systematic review and meta-analysis. *Journal of Affective Disorders*, 175, 424–440. doi:10.1016/j.jad.2015.01.050.
- Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, 67, 272–284. doi:10.1037/a0028015.
- Young, J. F., Mufson, L., & Davies, M. (2006). Efficacy of Interpersonal Psychotherapy-Adolescent Skills Training: An indicated preventive intervention for depression. *Journal of Child Psychology and Psychiatry*, 47, 1254–1262. doi:10.1111/j.1469-7610.2006.01667.x.
- Young, J. F., Mufson, L., & Gallop, R. (2010). Preventing depression: A randomized trial of interpersonal psychotherapy-adolescent skills training. *Depression and Anxiety*, 27, 426–433. doi:10.1002/da.20664.