



Published in final edited form as:

J Consult Clin Psychol. 2009 December ; 77(6): 1042–1054. doi:10.1037/a0017671.

A Meta-Analytic Review of the Penn Resiliency Program's Effect on Depressive Symptoms

Steven M. Brunwasser, M.S.,

University of Michigan

Jane E. Gillham, Ph.D., and

Swarthmore College and University of Pennsylvania

Eric S. Kim, B.A.

University of Michigan

Abstract

Objectives—The purpose of this review was to evaluate whether the Penn Resiliency Program (PRP), a group cognitive-behavioral intervention, is effective in targeting depressive symptoms in youth.

Data sources—We identified 17 controlled evaluations of PRP ($N = 2498$) measuring depressive symptoms via an online search of PsycInfo, Medline, ERIC, and ProQuest Dissertations and Theses, and by requesting data from PRP researchers.

Review methods—We combined effect sizes (ESs; Glass's d), using random effects models at post-intervention and two follow-up assessments.

Results—PRP participants reported fewer depressive symptoms at post-intervention and both follow-up assessments compared to youth receiving no intervention, with ESs ranging from 0.11 to 0.21. Limited data show no evidence that PRP is superior to active control conditions. Subgroup analyses showed that PRP's effects were significant at 1 or more follow-up assessments among studies using both targeted and universal approaches, when group leaders were research team members and community providers, among participants with both low and elevated baseline symptoms, and among boys and girls. Preliminary analyses suggest that PRP's effects on depressive disorders may be smaller than those reported in a larger meta-analysis of depression prevention programs for older adolescents and adults.

Conclusion—We found evidence that PRP significantly reduces depressive symptoms through at least 1 year post-intervention. Future PRP research should examine whether PRP's effects on depressive symptoms lead to clinically meaningful benefits for its participants, whether the program is cost-effective, whether CBT skills mediate program effects, and whether PRP is effective when delivered under real-world conditions.

Keywords

depression; prevention; youth; meta-analysis; Penn Resiliency Program

Introduction

Depression is one of the leading causes of disability worldwide (Murray & Lopez, 1997). Adolescence is a key time in the etiology of depression with rates increasing dramatically from the early to late teen years (Hankin, 2006). As many as 20–24% of youth have major depressive episodes by age 18 (Lewinsohn, Rhode, & Seeley, 1998). Elevated but sub-clinical levels of depressive symptoms are also common in adolescence (Roberts, Lewinsohn, & Seeley, 1991) and are associated with considerable impairment as well as increased risk for clinical depression (Gotlib, Lewinsohn, & Seeley, 1995). Widespread prevention efforts targeting adolescents may be our best hope at alleviating depression's enormous burden on society.

Researchers and mental health professionals have responded to this need by developing and testing prevention programs (see Sutton, 2007 for a recent review). These programs target a wide-range of risk factors, such as pessimistic cognitive styles, interpersonal difficulties, and family conflict. Most depression prevention programs are adapted from established psychotherapies for depression, such as cognitive-behavioral therapy (CBT) and interpersonal therapy (IPT).

In recent years, several research teams have published meta-analytic reviews of depression prevention programs (Cuijpers, van Straten, Smit, Mihalopoulos, & Beekman, 2008; Horowitz & Garber, 2006; Jané-Llopis, Hosman, Jenkins, & Anderson, 2003; Merry, McDowell, Hetrick, Bir, & Muller, 2004; Stice, Shaw, Bohon, Marti, & Rohde, 2009). These reviews have advanced prevention efforts considerably by allowing researchers to take stock of the existing literature and by raising important questions and recommendations for future prevention efforts. These meta-analyses indicate that youth who participate in depression prevention programs report lower levels of depressive symptoms than those who receive no intervention (Horowitz & Garber, 2006; Merry et al., 2004; Stice et al., 2009). In addition, participants in depression prevention programs are less likely to develop depressive disorders (Cuijpers et al., 2008).

The Penn Resiliency Program (PRP; Gillham, Reivich, & Jaycox, 2008) is one of the most widely researched depression prevention programs. PRP is a cognitive-behavioral group intervention designed for youth in late childhood and early adolescence (ages 10–14). Although typically a school-based program, researchers have evaluated PRP in other settings, including primary care clinics and juvenile detention centers. For a description of the intervention content see Gillham, Brunwasser, and Freres (2008). Findings from the initial efficacy study were promising as PRP prevented depressive symptoms through two years of follow-up and reduced the risk for clinically relevant symptoms (Gillham, Reivich, Jaycox, & Seligman, 1995). Since that initial study, however, PRP research findings have been inconsistent. The majority of studies evaluating PRP have found beneficial effects on depressive symptoms in either the overall sample or subgroups of participants. But at least

four studies found no significant effects (Gillham, et al., 2008). These conflicting results make it difficult to give an overall appraisal of the program's effectiveness. Few studies have evaluated PRP's effects on depressive disorders.

A priority for future PRP research is to determine whether PRP is likely to benefit youth if delivered on a wide scale, as intended. Large scale dissemination would require a considerable investment of time, effort, and finances. Such an investment is only justified if the existing data show promise. A meta-analytic review can help make this determination. We know of 17 controlled studies evaluating PRP's effects on depressive symptoms with more than 2000 participants in total. If a quantitative review of these studies detected no effect on depressive symptoms then it would be imprudent to continue evaluating the program in its current form or to disseminate the program broadly.

A second priority for PRP research is to understand the program's inconsistent effects. Research that identifies the contexts and subgroups in which PRP is most effective could guide future program development and implementation efforts. A meta-analysis can help to identify factors that moderate intervention effects. Detecting moderation, however, requires considerable statistical power as the analyst compares the strength of effects across subgroups of participants (Hedges & Pigott, 2004). Nearly half of the studies evaluating PRP have small samples ($N < 100$) limiting the power of meta-analytic analyses to detect moderators. Although it is unlikely that analyses comparing effects across subgroups would yield conclusive results at this time, there may be sufficient power to determine whether PRP's effects are significant within subgroups of interest.

It is plausible that PRP's inconsistencies are attributable to within- and between-study differences in participant characteristics. Prevention researchers who employ a *targeted* approach attempt to identify and recruit youth who are at increased risk for depression and, as such, in the greatest need of early intervention. Targeted intervention includes *selective* studies, in which participants have a known risk factor for the development of a disorder (e.g., parental depression), and *indicated* studies, in which participants evidence early symptoms of the disorder (e.g., sub-clinical depressive symptoms). In contrast, *universal* studies recruit all members of a specific population regardless of their level of risk. Depression prevention programs targeting at-risk youth have garnered more support than those delivered universally (Horowitz & Garber, 2006; Merry et al., 2004; Stice et al., 2009). PRP is one of the few depression prevention programs evaluated using both targeted and universal approaches. In this review, we examine the magnitude of PRP's effects in both targeted and universal studies.

Most studies of PRP have not found or examined potential moderators of intervention effects such as participants' sex or symptom level. A few studies have found that PRP's effects on depressive symptoms differ in boys and girls. At least one study (Gillham, Hamilton, Freres, Patton, & Gallop, 2006) found stronger effects for girls than boys, while other studies have found the opposite effect (e.g., Reivich, 1996). Some PRP studies have found that participants' pre-intervention levels of depressive symptoms moderated the intervention effects. For example, Gillham and colleagues found that PRP reduced the likelihood of receiving a diagnosis of depression or anxiety in participants with elevated but

not low baseline symptoms (Gillham, Hamilton, et al., 2006). In this review, we evaluate the magnitude of PRP's effects separately for boys and girls and for participants with elevated and low baseline symptoms.

A second possible source of inconsistency in PRP findings is within- and between-study differences in intervention provider characteristics. In some PRP studies, members of the research team (typically psychologists with extensive training in the cognitive-behavioral model, psychology graduate students, or advanced research assistants closely supervised by the program developers) led the intervention groups. In other studies, community providers, who would likely lead intervention groups if PRP were to be widely disseminated, led the intervention groups. Unlike research team members, community providers (e.g., school personnel or community mental health providers) typically do not have a direct interest in the research outcome. Gillham and colleagues expressed concern that PRP's inconsistent findings could be due partly to an attenuation of intervention effects when studies evaluate PRP under real-world conditions (i.e., effectiveness trials) as opposed to optimal research conditions (i.e., efficacy trials) (Gillham et al., 2008). A drop-off in intervention effects under real-world conditions would hamper dissemination efforts and limit PRP's utility. In this review, we evaluate PRP's effects among studies with both research team leaders and community providers.

The primary purpose of this meta-analysis was to aggregate data across all controlled studies to determine whether PRP participants have lower levels of depressive symptoms compared to youth who receive no intervention. Additionally, we conducted subgroup analyses to evaluate PRP's effects in different contexts. We expected to find more evidence for PRP's effects among targeted than universal studies. We expected to find more evidence for PRP's effects when research team members rather than community providers led intervention groups. We expected PRP's effects to be significant among both boys and girls and among participants with both low and elevated baseline symptoms. Because we lacked power to detect moderation, we did not focus on analyses comparing the strength of PRP's effects across subgroups. Although limited data were available, we ran preliminary analyses of PRP's effect on depressive disorders to determine whether the magnitude of PRP's effects is comparable to those reported in larger meta-analyses.

Method

Searching

We identified studies for this review using several methods. First, we conducted a search of several online databases: PsycInfo (1990–2009), Medline (1990–2009), ProQuest Dissertations & Theses (1990–2009), and ERIC (1990–2009). Search terms included all known names that have been used to describe PRP (*Penn Resiliency Program*, *Penn Prevention Program*, *Penn Optimism Program*, *Penn Program*, and *Depression Prevention Program*) and the names of the lead investigators of the PRP research team (*Gillham*, *Reivich*, *Jaycox*, *Shatté*, *Cardemil*, and *Seligman*). We limited searches to retrieve only articles describing empirical studies published no earlier than 1990, the year PRP was developed. The final online search date was February 28, 2009. Second, we cross-referenced the citation lists in each of the articles retrieved via the online search and reviewed the

citation lists of existing meta-analyses to ensure we uncovered all PRP studies included in these reviews. Finally, we consulted a database maintained by the program developers since February 2003 that records all research-related requests for the PRP program materials. We contacted all researchers who had requested the PRP program materials and asked them to provide data from their studies and to complete a survey asking for details about the study design, participants, group leaders, and intervention delivery.

Selection

Studies included in the review (1) compared PRP to a control condition, (2) evaluated PRP's effect on depressive symptoms, and (3) reported data both before the intervention began (baseline) and at one or more post-intervention assessments. No studies were excluded due to sub-optimal research methods (e.g., non-random assignment), however, we report intervention effects both including and excluding non-randomized studies. The review includes data from both published and unpublished studies.

The first author reviewed the abstracts of all articles retrieved via the online database search and obtained the full-text for each article that mentioned PRP by name or described a cognitive-behavioral intervention for youth. Both the first and second authors reviewed the study descriptions provided by the researchers who responded to our request for data to determine if their studies met inclusion criteria. The final determination of which studies were to be included in the review was made by the consensus of the first and third authors.

Data Abstraction

The first author coded all study data into an MS Access database and wrote algorithms to calculate ESs. A trained undergraduate research assistant (the third author) served as an independent coder and reentered all data. When data needed to compute ESs were not available in study manuscripts, we contacted the manuscript author(s). In all cases, the authors provided the necessary data to calculate effects on depressive symptoms. We also coded three dichotomous dummy variables representing between-study subgroups of interest: *Condition Assignment* (random vs. non-random), *Participant Risk Status* (universal vs. targeted), and *Group Leader Type* (research team members vs. community providers).

We were also interested in evaluating PRP's effects across two within-study factors: *Sex* (girls vs. boys) and *Symptom Level* (participants with low vs. elevated baseline symptoms). Few PRP studies have reported summary statistics for these subgroups. However, we had access to full data sets for nine studies included in this review, allowing us to calculate ESs by *Sex* and *Symptom Level*. A total of 10 studies provided sufficient data to calculate separate ESs for boys and girls and 9 studies provided sufficient data to calculate separate ESs for participants with low and elevated baseline symptoms. We classified participants as having either low or elevated baseline symptoms based on a CDI cutoff score of 13, a recommended cutoff score (Kovacs, 2001). One study (Roberts, Kane, Thomson, Bishop, & Hart, 2003) reported separate data for participants with low and elevated baseline symptoms based on a CDI cutoff of 15. We chose to include data from this study in the subgroup analyses because the cutoff score was close to the one selected for the other studies.

We then exported the data into the Comprehensive Meta-Analysis Version 2.2.046 software (CMA; Biostat, Englewood New Jersey) to conduct analyses. In addition to coding ES data and moderators, the first and third authors coded information related to the research design, participant demographics, and intervention delivery for each study. We conducted no formal evaluation of study design quality but provide detailed information about each study in online supplemental tables.

Effects on Depressive Symptoms

Power Analysis—In order to gauge our ability to detect effects on depressive symptoms, we conducted power analyses following the procedures described by Hedges & Pigott (2001). We calculated our power to detect an effect size of 0.20 ($\alpha = .05$), an effect that is of a magnitude similar to those reported by recent meta-analytic reviews of youth depression prevention programs (see Supplemental Table 5).

Calculating Effect Sizes—We calculated ES estimates (i.e., standardized mean difference scores) for depressive symptoms by dividing the difference in the control group and PRP group means by the control group's standard deviation (Glass's d ; Glass, McGaw, & Smith, 1981): $d = (X_{\text{control}} - X_{\text{PRP}}) / SD_{\text{Control}}$. Positive d values indicate fewer depressive symptoms in PRP groups compared to control groups. Standardized mean difference scores based on small samples tend to be upwardly biased (Hedges, 1981). We applied Hedges's (1981) correction to all d estimates to create an unbiased ES estimate: $d^U = d \times [1 - (3/4df - 1)]$. Hedges's correction reduces overestimation of the ES in small studies but has a negligible effect on ES estimates in large studies. When the standard deviation in the denominator of the ES is based on 50 degrees of freedom or more, d and d^U are nearly identical (Hedges, 1981).

Studies evaluating PRP have differed in their length of follow-up. We limited analyses to the three most commonly reported assessments (post-intervention, 6- to 8-month follow-up, and 12-month follow-up) and calculated separate ESs for each. When studies measured the same outcome variable with more than one instrument, we computed an average ES estimate across the different instruments so that no study provided multiple ESs at a given assessment. When studies had more than one PRP condition (e.g., an adolescent-only PRP group and an adolescent + parent PRP group), we pooled the means and standard deviations of the different PRP conditions in order to calculate one ES. When studies compared PRP to both a no-intervention control condition and an active control condition, we calculated separate ESs comparing PRP to both control conditions.

Assessing Heterogeneity—We used Q tests at all assessments to determine whether there were any significant violations of homogeneity in the ES distributions. We also evaluated the proportion of heterogeneity between studies using the I^2 statistic because homogeneity tests tend to be underpowered (Higgins, Thompson, Deeks, & Altman, 2003).

Combining Effect Sizes—We used random effects models when combining ESs across studies. Fixed effects models assume that between-study differences are due to sampling error alone (Cooper & Hedges, 1994). In contrast, random effects models assume that, in

addition to sampling error, there are other sources of between-study variability. Random effects models add a separate variance term (ν_{ρ}) to account for non-sampling error. This results in larger ES confidence intervals (Lipsey & Wilson, 2001). The assumptions of random effects models seemed more appropriate for this review given that there are considerable methodological differences across PRP studies. We followed procedures recommended by Lipsey & Wilson (2001) when calculating mean ESs. The unbiased standardized mean difference score (d^U) for each study was weighted by its inverse variance (ω): $\omega = 1/(SE^2 + \nu_{\rho})$. The weighted ESs were then added and divided by the sum of the inverse variance weights across all studies. This produced a mean ES (d_+) for each assessment.

Converting Effect Sizes—Although standardized mean difference scores are statistically intuitive, they do not lend themselves readily to clinical interpretation (Acion, Peterson, Temple, & Arndt, 2006). To facilitate comprehension of ESs, we converted standardized mean difference scores into more easily interpretable metrics. First we converted the mean ESs on the CDI from standard deviation units to the CDI's scoring metric. We did this by multiplying the mean ES for all studies that used the CDI ($k = 16$) by the pooled CDI standard deviation across the control groups. This product represents the average benefit of PRP in the CDI metric (Lipsey & Wilson, 2001). A product of 0.50, for example, means that PRP groups scored, on average, half a point lower than the control groups on the CDI.

Second, we converted ESs into estimates of the *probability of superiority* (PS). The PS score is an estimate of the probability that a randomly selected PRP participant had a favorable outcome (i.e., lower depressive symptoms) compared to a randomly selected control participant. A PS score of 0.50 indicates that there is 50% chance that a randomly selected PRP participant has a better score than a randomly selected control participant (i.e., no intervention effect). Scores ranging from 0.51 to 1.00 indicate preferable outcome for PRP participants whereas scores from 0.00 to 0.49 indicate a preferable outcome for control participants (Grissom, 2005). When full data sets were accessible, we calculated PS by dividing the Mann-Whitney U statistic by the product of the sample sizes for the PRP and control conditions: $PS = U/mn$. We used an approximate conversion method when there were insufficient data to calculate a U statistic: $PS = \Phi(d^U / \sqrt{2})$, where Φ is the normal cumulative distribution function. The non-parametric Mann-Whitney U method is preferable because it does not assume that both groups (in this instance, PRP and control) have normal distributions (Acion, et al., 2006); an assumption that is untenable when evaluating depressive symptoms in non-clinical samples.

One can easily convert PS scores into a number needed to treat (NNT) score: $NNT = 1/[(2 \times PS) - 1]$. NNT, in this circumstance, represents the approximate number of youth who need to receive PRP, rather than the control condition, to yield one superior outcome. A superior outcome is when a randomly selected PRP participant has a better depressive symptom score than a randomly selected control participant (Kraemer & Kupfer, 2005). We provide PS and NNT scores for each study in Supplemental Table 7.

Sensitivity analyses—We conducted several forms of sensitivity analysis to determine whether effects on depressive symptoms were robust. We evaluated the influence of each individual study on the mean ESs with the *one-study-removed* procedure in CMA. This is an iterative procedure in which mean ESs and confidence intervals are repeatedly recalculated excluding one study at a time from the analysis. This procedure allows one to determine whether any individual study was influential enough to alter the decision about whether to reject the null hypothesis (i.e., $d_+ = 0$).

Studies with null findings are less likely to be published and thus, more likely to go undiscovered by reviewers. The “file-drawer problem” causes systematic bias often leading to an overestimation of effects in meta-analysis (Rosenthal, 1979). We assessed for the existence and impact of publication bias using two procedures. First, we examined funnel plots and normal-quantile plots of study ESs at each assessment. These plots allow the analyst to detect gaps in the ES distribution that could be indicative of publication bias. If publication bias were not a concern, one would expect the distribution of study ESs to be normal (Light, Singer, & Willett, 1994; Wang & Bushman, 1998). We then recalculated mean ESs adjusting for the possible effect of undiscovered studies using *trim-and-fill* analyses. *Trim-and-fill* makes the distribution of ESs on a funnel plot normal by eliminating outlying ESs and imputing ES estimates for hypothetically missing studies. The mean ES is then recalculated with the imputed studies. If the mean ES remains significant, one can have increased confidence that missing studies would not have altered the decision about whether to reject the null hypothesis (Duval & Tweedie, 2000).

Our inclusion of non-randomized studies could be another source of bias. Random assignment to study conditions ensures that baseline between-group differences are due to chance. Non-randomization could add systematic error to ES estimates as differences at the post-intervention data points could reflect baseline differences rather than intervention effects. To ensure that study effects were not driven by non-randomized studies, we reran all primary outcome analyses excluding non-randomized studies.

Subgroup & Moderator Analyses—We used Q tests to evaluate whether our subgroup variables (*Participant Risk Status, Group Leader Type, Symptom Level, and Sex*) accounted for systematic variance in PRP’s effects. Q tests are akin to ANOVA in that they compare within- and between-group variance (using a χ^2 test statistic) to determine whether variability between groups exceeds chance expectation (Lipsey & Wilson, 2001). When evaluating moderators, we used mixed effects modeling which assumes that there are both systematic and non-systematic sources of heterogeneity in ES estimates (Lipsey & Wilson, 2001). Mixed effects models use random effects modeling when aggregating ESs within subgroups and fixed effects modeling when aggregating across subgroups (Overton, 1998).

Effects on Depressive Disorders

As noted, few studies have evaluated PRP’s effects on diagnostic outcomes limiting statistical power to detect effects on depressive disorders. However, we chose to run preliminary analyses with the available diagnostic data in an effort to determine whether the

effects in the existing PRP studies are comparable to those reported in a recent, larger meta-analysis of depression prevention programs (Cuijpers et al., 2008).

Following the example of Cuijpers and colleagues, we evaluated PRP's effects on depressive disorders using both relative risk and relative incidence analyses (Cuijpers et al., 2008). Relative risk analyses compare the proportion of participants in each condition who experience the outcome of interest (i.e., depressive disorders) over the follow-up. For each study providing diagnostic data, we calculated risk ratios (RRs) by dividing the PRP group risk (i.e., the percentage of PRP participants who received a depression diagnosis) by the control group risk. Risk ratios smaller than 1.00 indicate a beneficial effect of PRP whereas scores greater than 1.00 indicate a benefit for the control condition. We also calculated the number needed to treat (NNT) which represents the number of participants that would need to receive the intervention in order to prevent one case of depression. The NNT is calculated by taking the inverse of the difference in risk between the control and PRP conditions: $NNT = 1/(Risk_{Control} - Risk_{PRP})$ (Woodward, 2005).

The relative risk approach assumes that all participants completed an equal number of diagnostic assessments covering an equal amount of time. This was an unsound assumption in this review because individual studies differed in their length of follow-up and many participants had incomplete data. We computed a person-years (PY) score for each participant in order to account for the discrepancy in the number of diagnostic assessments completed. PY scores reflect the total number of years during the follow-up that the person went without receiving a depression diagnosis. For example, if a participant completed three assessments each covering a 6-month span without receiving a diagnosis, that person contributed 1.5 PYs to the analysis. Once a participant met criteria for a depressive disorder, he/she stopped contributing PYs. We then calculated the incidence of depression in both the PRP and control groups by dividing the total number of participants receiving a depression diagnosis at some point during the follow-up by the total number of PYs across participants. We then calculated the incidence rate ratio (IRR) by dividing the PRP group's incidence rate by the control group's incidence rate. IRRs less than 1.00 reflect a benefit of PRP.

We computed both a mean IRR (IRR_{+}) as well as a mean RR (RR_{+}) using random effects models. Additionally, we evaluated whether PRP's effect on depressive disorders among two subgroup variables: *Sex* (boys and girls) and *Symptom Level* (low vs. elevated baseline depressive symptoms based on a CDI cutoff score of 13).

Results

Study Flow

Our online database search yielded 519 manuscripts, 44 of which either identified PRP by name or described a cognitive-behavioral prevention program for youth in the abstract. We excluded 16 of these studies, after reviewing the full-text of the articles, because they did not describe evaluations of PRP. An additional six articles describing PRP were eliminated because they either did not report depression data ($k = 2$) or did not have a control condition ($k = 4$). The remaining 22 manuscripts reported data from 15 separate evaluations of PRP that met our inclusion criteria (see Supplemental Table 1). We contacted 19 researchers who

requested the PRP manuals for research purposes and received responses from 15. Most of these researchers ($n = 9$) indicated that they had not yet conducted evaluations of PRP. Of the six studies evaluating PRP, we excluded four because they either did not assess depressive symptoms ($k = 2$) or did not have a control condition ($k = 2$). The remaining two studies met our inclusion criteria. Thus, a total of 17 evaluations of PRP were included in this review (see Figure 1).

Study Characteristics

A total of 2498 youth participated in the 17 PRP studies included in the review. Participants ranged in age from 8 to 18 years old. Most studies included in this review used some form of random condition assignment ($k = 14$; $n = 2281$), either randomizing at the participant, classroom, or school level. Three studies provided data only at baseline and immediate post-intervention assessments, while others evaluated intervention effects as late as three years post-intervention. Most studies included in the review used a targeted ($k = 11$; $n = 1408$) rather than a universal ($k = 6$; $n = 1090$) intervention approach. An equal number of studies had research team members ($k = 8$; $n = 521$) and community providers ($k = 8$; $n = 1884$) leading intervention groups but the studies with community providers tended to be much larger. Community providers included school staff (i.e., teachers and counselors), learning mentors, and child mental health professionals working for a managed care organization. In a few studies, school staff led the vast majority of intervention groups with research team members leading a small number of groups; these studies were coded as having community providers as group leaders. One study (Reivich, 1996) had an equal number of researchers and school staff leading intervention groups and was excluded from subgroup analyses of researchers and community providers.

Four studies compared PRP to both a no-intervention control condition and an active control condition. Two of these studies (Gillham, Reivich, Freres, et al., 2007; Reivich, 1996) compared PRP to the Penn Enhancement Program (PEP), an alternative intervention designed specifically to mimic the “non-cognitive modes of action” (such as adult attention, group cohesion, and the discussion of day-to-day problems and feelings) that likely contribute to PRP’s effects (Reivich, 1996, p. 23). PEP includes psychoeducation and non-cognitive skill building exercises (e.g., techniques for goal setting, communication, and resisting peer pressure) designed to be relevant to youth with depressive symptoms (Reivich, 1996; Shatté, 1996). Two studies (Pattison & Lynd-Stevenson, 2001; Wass, 2008) compared PRP to conditions designed to control for social contact and group cohesion (see Supplemental Table 10).

All but one of the 17 studies included in this review measured depressive symptoms with the Children’s Depression Inventory (CDI; Kovacs, 2001). Two studies measured depressive symptoms with both the CDI and the Reynolds Adolescent Depression Scale (Reynolds, 1986). One study measured depressive symptoms with the Depression Self-Rating Scale (Birlerson, 1981). Only three studies evaluated PRP’s effects on depressive disorders. Two of these studies assessed for depressive disorders using standardized diagnostic interviews: the Children’s Depression Rating Scale—Revised (Poznanski & Mokros, 1996) and the Diagnostic Interview Schedule for Children, Version IV (Shaffer, Fisher, Lucas, Dulcan, &

Schwab-Stone, 2000). Participants completed these interviews at regular intervals during the studies. The third study evaluated PRP's effects on depressive disorders using HMO computerized medical records (Gillham, Hamilton, et al., 2006).

Coder Agreement

The coders achieved a high level of reliability for the continuous ES data ($\alpha_s > .90$) and achieved perfect agreement in coding both *Condition Assignment* and *Participant Risk Status* ($\kappa_s = 1.00$). The raters had a reliability rating of $\kappa = .79$ when coding *Group Leader Type* (15 agreements and 2 discrepancies). The first and second authors resolved all coding discrepancies.

Power Analyses

We had a considerable amount of statistical power (0.88 to 0.98) to detect an effect of 0.20 in our analyses with the overall sample. The power of subgroup analyses was greater than 0.50 except among the subgroup of participants with elevated symptoms and among the subgroup of studies with research team group leaders (see Supplementary Table 5).

Heterogeneity Assessment

There was no evidence that the amount of variability between study ESs exceeded chance expectation at any assessment, $\chi^2_{\text{post}}(16) = 21.14, p = .17, \chi^2_{6-8\text{-month}}(12) = 12.54, p = .40$, and $\chi^2_{12\text{-month}}(9) = 6.20, p = .72$. The proportion of heterogeneity between studies was less than 25% (which is considered low) at all assessments, $I^2_{\text{post}} = 24.30, I^2_{6-8\text{-month}} = 4.28$, and $I^2_{12\text{-month}} = 0.00$ (Higgins et al., 2003). See Supplemental Table 6 for details on heterogeneity analyses.

Effects on Depressive Symptoms

The mean ES comparing PRP and no-intervention control conditions at post-intervention was significant (i.e., significantly different than zero), $d_+ = 0.11$ (95% CI = 0.01, 0.20). PRP groups had fewer depressive symptoms than control groups in 14 of 17 studies with ESs ranging from -0.61 to 0.59 . On average, PRP groups scored 0.86 points lower on the CDI (indicating fewer depressive symptoms) than control groups, and PS score ranged from 0.33 to 0.66. The mean ES was also significant at the 6- to 8-month follow-up, $d_+ = 0.21$ (95% CI = 0.11, 0.31). ESs ranged from -0.06 to 0.69 , and PRP groups had fewer depressive symptoms than control groups in 12 of 13 studies. The average benefit of PRP was 1.75 points on the CDI at the 6- to 8-month follow-up, and PS scores ranged from 0.48 to 0.69. The mean ES remained significant at 12-month follow-up, $d_+ = 0.20$ (95% CI = 0.09, 0.32). ESs ranged from -0.10 to 0.61 , and PRP groups had fewer depressive symptoms than control groups in 9 of 10 studies. PRP groups scored, on average, 1.56 points lower on the CDI than control groups at 12-month follow-up, and PS scores ranged from 0.47 to 0.67. See Table 1 for a summary of ESs at each assessment.

The mean ES comparing PRP to active control conditions was not significant at either post-intervention or 6- to 8-month follow-up, $d_{+\text{post}} = -0.02$ (95% CI = $-0.19, 0.14$) and $d_{+6-8\text{month}} = 0.00$ (95% CI = $-0.18, 0.19$). PRP groups had lower mean depressive symptom scores than the active control conditions in only one of four studies at post-intervention and

in only one of three studies at 6-to 8-month follow-up (see Supplemental Table 10). Only one study compared PRP to an active control condition at the 12-month follow-up precluding meta-analytic analyses. The active control conditions had lower mean levels of symptoms than no-intervention control conditions in all four studies reporting data at post-intervention, $d_+ = 0.10$ (95% CI = $-0.07, 0.26$), and in all three studies reporting data at the 6- to 8-month assessment, $d_+ = 0.14$ (95% CI = $-0.05, 0.33$). These effects were not significant but were based on limited data ($N_{\text{post}} = 568$ and $N_{6-8\text{-Month}} = 428$).

Sensitivity analyses—Because the mean ESs comparing PRP and active conditions were not significant, we limited sensitivity analyses to effects of PRP compared to no-intervention control conditions. Findings from the sensitivity analyses differed considerably between post-intervention and the two long-term follow-ups. At post intervention, the *one-study-removed* procedure showed that 6 of 17 studies carried enough weight that their removal from the analysis would have made the mean ES non-significant. Additionally, the post-intervention mean ES became non-significant when adjusting for publication bias using the *trim-and-fill* procedure, $d_+ = 0.09$ (95% CI = $-0.01, 0.19$), and when removing studies using a non-randomized design, $d_+ = 0.09$ (95% CI = $-0.02, 0.19$). Therefore, the post-intervention effect, while significant, is precarious and warrants cautious interpretation. In contrast, there was considerable evidence that the long-term follow-up effects were robust. No single study when removed from analyses carried enough weight to nullify the mean ES at either follow-up assessment. Additionally the mean ESs remained significant after adjusting for possible publication bias with the *trim-and-fill* procedure, $d_{+6-8\text{-month}} = 0.17$ (95% CI = $0.07, 0.28$) and $d_{+12\text{-month}} = 0.17$ (95% CI = $0.06, 0.28$), and when excluding non-randomized studies, $d_{+6-8\text{-month}} = 0.20$ (95% CI = $0.09, 0.31$) and $d_{+12\text{-month}} = 0.18$ (95% CI = $0.07, 0.31$).

Subgroup Analyses: Between-study factors—This review had limited power to detect significant moderation, and heterogeneity analyses showed that there was little between-study variation to capture in moderator analyses. None of our hypothesized moderators accounted for a significant amount of heterogeneity in ESs. Therefore, we focus on analyses evaluating whether PRP's effects were significant in subgroups of interest. Moderation statistics (between-group Q statistics) are available in Supplemental Table 6.

The mean ES for targeted studies was significant at all three assessments: $d_{+\text{post}} = 0.14$ (95% CI = $0.01, 0.26$); $d_{+6-8\text{-month}} = 0.23$ (95% CI = $0.11, 0.36$); $d_{+12\text{-month}} = 0.22$ (95% CI = $0.06, 0.38$). The mean ES among universal studies was significant at the 12-month follow-up, $d_+ = 0.19$ (95% CI = $0.01, 0.37$), but not at post-intervention, $d_+ = 0.06$ (95% CI = $-0.10, 0.23$), or the 6- to 8-month follow-up, $d_+ = 0.15$ (95% CI = $-0.02, 0.33$). The effects among both research team leaders and community providers were non-significant at post-intervention: $d_+ = 0.20$ (95% CI: $-0.02, 0.41$) and $d_+ = 0.08$ (95% CI: $-0.04, 0.19$), respectively. The mean ESs for both research team and community leaders were significant at the 6–8-month assessment, however: $d_+ = 0.29$ (95% CI: $0.06, 0.53$) and $d_+ = 0.17$ (95% CI: $0.06, 0.28$), respectively. The mean ESs for both research team and community leaders remained significant at 12-month follow-up: $d_+ = 0.31$ (95% CI: $0.03, 0.60$) and $d_+ = 0.18$ (95% CI: $0.05, 0.32$), respectively (see Supplemental Table 7).

Subgroup Analyses: Within-study factors—PRP's effect among girls was significant at 6–8-month follow-up, $d_+ = 0.19$ (95% CI: 0.02, 0.35), but not at post-intervention, $d_+ = 0.06$ (95% CI: –0.11, 0.22) or 12-month follow-up, $d_+ = 0.16$ (95% CI: –0.01, 0.32). PRP's effects among boys were significant at both follow-up assessments, $d_{+6-8\text{-Month}} = 0.21$ (95% CI: 0.05, 0.37) and $d_{+12\text{Month}} = 0.25$ (95% CI: 0.08, 0.41), but were not significant at post-intervention, $d_+ = 0.05$ (95% CI: –0.12, 0.22). PRP's effects among low symptom participants were significant at all assessments: $d_{+\text{post}} = 0.13$ (95% CI = 0.02, 0.24); $d_{+6-8\text{-month}} = 0.15$ (95% CI = 0.01, 0.29); $d_{+12\text{-month}} = 0.19$ (95% CI = 0.04, 0.34). Effects among participants with elevated symptoms were significant at both follow-up assessments, $d_{+6-8\text{-month}} = 0.28$ (95% CI = 0.03, 0.53) and $d_{+12\text{-month}} = 0.27$ (95% CI = 0.04, 0.51), but not at post-intervention, $d_{+\text{post}} = 0.18$ (95% CI = –0.03, 0.39) (see Supplemental Table 9).

Depressive Disorders

The mean IRR comparing PRP and no-intervention control conditions was 0.89 (95% CI: 0.64, 1.24), indicating that PRP participants were approximately 11% less likely to receive a depression diagnosis. Individual study IRRs ranged from 0.80 to 1.10. The mean RR was 0.90 (95% CI: 0.66, 1.23) indicating a risk reduction of 10% in the PRP group. Neither of these effects represents a significant benefit of PRP. Overall, 75 out of 622 PRP participants (totaling 1238 PYs) met criteria for a depressive disorder as compared to 68 out of 470 control group participants (totaling 920 PYs). The NNT across all three studies was equal to 41 (see Supplemental Table 11).

PRP did not significantly reduce the risk for depressive disorders among any subgroups examined. However, preliminary analyses suggest that boys and participants with elevated symptoms may benefit from PRP more than girls and low symptom participants. Among boys, the mean IRR was 0.74 (95% CI: 0.45, 1.24), representing a 26% reduction in incidence in the PRP group, compared to a mean IRR among girls of 1.02 (95% CI: 0.65, 1.59). PRP participants with elevated symptoms were 16% less likely to have a diagnosis [$\text{IRR}_+ = 0.84$, (95% CI: 0.52, 1.36)] compared to no-intervention control participants with elevated symptoms, while there was practically no benefit for low symptom PRP participants [$\text{IRR}_+ = 0.94$, (95% CI: 0.58, 1.51)]. The NNT among boys and participants with elevated symptoms was 23 and 16, respectively, compared to 238 and 65 among girls and low symptom participants (see Supplemental Table 12).

Discussion

Effects on Depressive Symptoms

The primary goal of this meta-analysis was to determine whether PRP is effective in targeting depressive symptoms. We found that youth who participate in PRP report reliably lower levels of depressive symptoms through 12 months of follow-up compared to youth who receive no intervention. The effects are modest (ranging from 0.11 to 0.21) but of a similar magnitude to those reported in larger meta-analyses of depression prevention programs (Horowitz & Garber, 2006; Jané-Llopis et al, 2003; Merry et al., 2004; Stice et al., 2009). (Direct comparisons with other depression prevention meta-analyses should be made cautiously given that there are important methodological differences between the studies.)

On average, PRP groups scored between 0.86 and 1.75 points lower on the CDI than no-intervention control groups. A single point on the CDI is indicative of a change in the frequency or intensity of a depressive symptom.

It is unclear at this time why PRP's effects became more robust at the follow-up assessments than at post-intervention. We considered the possibility that control participants had an increase in depressive symptoms in the first year following the study creating more room for an intervention effect. To evaluate this possibility, we calculated mean depressive symptom scores at each assessment across all studies using the CDI, weighting each study's mean symptom score by its sample size. Contrary to expectation, the mean control group CDI score tended to decrease over the follow-up (from $M = 9.39$ at post to $M = 8.80$ at 12-month follow-up). In 7 of the 8 studies reporting 12-month data, the control group reported decreases in symptoms from post-intervention to 12-month follow-up. The strengthening of PRP's effects cannot be attributed to an increase in control group symptoms. We also considered the possibility that studies with small post-intervention effects were less likely to collect follow-up data leading to overestimates of mean ESs at follow-up assessments. This seems unlikely given that two of the three studies (Tellier, 1998; Wass, 2008) with only post-intervention data available had larger than average effects (0.39 and 0.51, respectively). It is also possible that PRP's effects truly grow stronger over time. It may take time before students begin applying the program skills in their everyday lives. Future research should examine the relationship between participants' use of the PRP skills and their depressive symptoms over time.

The limited data available show no evidence that PRP is superior to active control conditions at either post-intervention or 6–8-month follow-up. The dearth of statistical power in these analyses limits our ability to draw firm conclusions. However, the mean ES was not even in the expected direction at either post-intervention or 6–8-month follow-up suggesting that PRP is not superior to active control conditions. Future research should continue to compare PRP to active control conditions in terms of mental health outcomes and in terms of cost and ease of delivery.

Secondary Analyses

We also conducted moderator and subgroup analyses evaluating whether participant and group leader characteristics influence PRP's effects. There was no evidence that any subgroup variables accounted for a significant amount of heterogeneity. Our ability to detect moderation was limited due to the relatively small sample of studies ($k = 17$), many of which were underpowered. However, it is important to note that heterogeneity levels were low (particularly at the follow-up assessments). This could mean that the between-study differences have little impact on PRP's effects. As PRP research accumulates, meta-analysts should continue to evaluate whether contextual factors moderate PRP's effects.

Participant characteristics—PRP's effects tended to be larger (though not significantly) at all three assessments when delivered to targeted samples than when delivered universally. This is not surprising given that there is typically more room for an effect in targeted studies. Consistent with two previous meta-analyses (Horowitz & Garber, 2006; Merry et al., 2004)

we found no evidence for PRP's effectiveness in universal studies at the post-intervention or 6–8-month follow-up assessments. We did find a significant effect of universal studies on depressive symptoms at the 12-month follow-up ($d_+ = 0.19$), however. This finding is consistent with findings from a recent meta-analysis of depression prevention programs in which effects of universally delivered interventions were not significant at post-intervention but were significant (though smaller than effects of targeted studies) across the long-term follow-up (Stice et al., 2009). Several prevention researchers have suggested that further research investigating universal prevention may not be warranted (e.g., Spence & Shortt, 2007; Stice et al., 2009). However, given their significant longer-term effects and their potential to reach large numbers of youth, we believe it is important to continue efforts to develop and evaluate universal depression prevention programs.

We found evidence for PRP's effectiveness among both boys and girls. The mean ES among boys was significant at both follow-up assessments, while the mean ES for girls was only significant at 6–8-month follow-up. It is noteworthy that the range in ESs among boys and girls was considerable. In two studies, there were particularly large discrepancies in the effects among boys and girls. The ESs for boys in the Reivich (1996) study were relatively large (ranging 0.35 to 0.61) while the effects for girls were remarkably poor (ranging from -0.39 to 0.06). Conversely, a different study (Gillham, Hamilton, et al., 2006) yielded consistently positive effects for girls (ranging from 0.21, to 0.34) and poorer effects for boys (ranging from -0.33 to 0.16). These findings suggest that contextual factors (e.g., the intervention setting or group leader characteristics) may influence PRP's effects on boys and girls differently. For example, it could be that having single-sex PRP groups is beneficial for girls but not so for boys. PRP group leaders have noted in supervision that girls seem more engaged in the intervention and feel more comfortable talking about sensitive issues when the group is predominantly or entirely female. Chaplin and colleagues found that girls in single-sex groups attended more PRP sessions and had lower hopelessness scores than girls in coed groups, although both all-girls and co-ed PRP led to similar improvements in depressive symptoms relative to a no-intervention control (Chaplin et al., 2006). A study evaluating the influence of group characteristics (such as the gender composition of groups) on PRP outcomes is underway.

Group leader characteristics—The mean ESs for studies with research team leaders tended to be larger (although not significantly) than those among studies with community leaders at all three assessments; however, the mean ESs were significant at both follow-ups regardless of whether intervention providers were primarily members of the research team or community leaders. As more studies of PRP are conducted, it will be important to revisit the question of whether there is a drop-off in intervention effects when real-world personnel lead intervention groups. It is encouraging that PRP's effects were significant with community leaders as effective dissemination is contingent upon PRP's success when led by real-world personnel.

Effects on Depressive Disorders

PRP did not have a significant effect on diagnoses of depression. Given that only three studies of PRP measured diagnostic outcomes we did not expect to have enough statistical

power to detect a significant intervention effect. PRP participants were only 11% less likely than controls to receive a diagnosis and 41 participants are needed to prevent one case of depression. Our preliminary analyses suggest there may be diagnostic benefits for boys and participants with elevated symptoms but there is no evidence of benefit for girls or low symptom participants.

Very few studies of depression prevention programs for youth have measured effects on depression diagnoses. A recent meta-analysis of depression prevention studies for adolescents and adults found that participants in prevention programs were 23% less likely than controls to be diagnosed with depression and that 21 participants needed to receive the intervention to prevent one case of depression (Cuijpers et al., 2008). PRP's effects on diagnosis appear to be about half this size. There are several possible explanations for this discrepancy. It is possible that PRP's effects on depressive symptoms do not translate into prevention of diagnoses. Alternatively, the discrepancies may reflect differences in participants' ages. The vast majority of studies that have examined the prevention of disorder include participants in late adolescence through adulthood when depression rates are high. In contrast, PRP targets early adolescents, who are far less likely to have clinical depression and who may have more difficulty learning and applying cognitive-behavioral skills. A third possibility is that differences in effects reflect differences in risk status. On average, participants in the PRP studies that examined diagnoses scored 9.9 on the baseline CDI (which is between the 57th and 69th percentile depending on participant age and sex; Kovacs, 2001), while most other studies examining prevention of depressive disorders have selected participants at substantially elevated risk.

Questions & Recommendations for Future Research

This meta-analysis indicates that PRP participants have reliably lower levels of depressive symptoms compared to participants who receive no intervention, and these effects endure for at least 12-months after the intervention. However, this review leaves us with more questions than answers. PRP was developed with the intention of widespread implementation leading to a considerable decrease in the burden of depression. Clearly, this lofty goal is far from accomplished. Future research should address the following questions.

(1) Are PRP's effects meaningful?—The most important objective for future PRP research will be to demonstrate that PRP's effects have practical significance. We propose a broad definition for what constitutes a meaningful intervention effect as one that leads to improvements in the emotional health or functioning of the participants, their family members, and/or peers. There are many ways in which PRP *could* produce meaningful benefits including, but not limited to, the following: (a) preventing, delaying, or lessening the intensity or duration of future psychological disorders; (b) eliminating or ameliorating the distress and impairment associated with sub-clinical symptoms of depression, anxiety, or externalizing problems; (c) improving interpersonal relationships with peers and/or family members; (d) increasing awareness of youth depression among participants, teachers, and guardians, and improving their ability to respond effectively; and (e) improving parental well-being and parenting practices (which is the goal of the parent intervention component).

At this time, it is unclear whether PRP yields these benefits. There is no evidence at this time that PRP satisfies our criterion (a) as effects on depressive disorders were not significant. PRP is closest to meeting criterion (b). PRP has enduring effects on depressive symptoms but it is unclear whether an average reduction in symptoms by one-fifth of a standard deviation translates into practical benefits for youth. PRP's effects are small by many intervention researchers' standards (e.g., Weisz, Donenberg, Han, & Weiss, 1995). But the meaningfulness of an effect is not simply a function of its magnitude (Prentice & Miller, 1992). Many widely-accepted interventions yield small effects, comparable to PRP's effects on depressive symptoms (Meyer et al., 2001). The important question is whether PRP's small effect on depressive symptoms is a mediator of practical benefits for youth (e.g., decreased risk for depressive disorders, improved adaptive functioning, quality of life, etc.). Future PRP studies should include outcome measures that better lend themselves to clinical interpretation.

In the short term, effects among youth with elevated depressive symptoms are likely more meaningful than effects among students who already have low levels of symptoms. This review suggests that PRP is effective in reducing symptoms among students with elevated baseline symptoms. But PRP is not intended to be a short-term treatment program; rather it is intended to impart lasting skills that will reduce the risk for depression as youth enter late adolescence and early adulthood. Unfortunately, few PRP studies have been able to follow youth into this period of heightened risk. Extending follow-up periods would be difficult due to increased costs and attrition, but doing so would improve our ability to gauge PRP's potential benefits. It is noteworthy that, although screening instruments can be effective in identifying youth at increased risk for depression, many (and perhaps more) youth who score low on these instruments at a screening or baseline assessment will ultimately develop significant levels of depression (Gillham, 2003). Thus, we feel that, in the long-term, PRP's effects among low symptom youth could be just as meaningful as its effects among participants with elevated symptoms.

It is likely that PRP's effects extend beyond depression. The program is based on CBT skills that are used in the treatment of a variety of psychological disorders (Butler, Chapman, Forman, & Beck, 2006). Given their high levels of comorbidity with depression, the PRP program developers included content specifically targeting anxiety and externalizing symptoms. Few PRP studies have evaluated these outcomes, but there is some evidence that PRP can improve anxiety and externalizing symptoms (Gillham, Reivich, et al., 2006; Jaycox, Reivich, Gillham, & Seligman, 1994; Roberts, Kane, Bishop, Matthews, & Thomson, 2004). Research that evaluates PRP's effects on anxiety, behavioral problems, and other outcomes can lead to better estimates of the program's true impact.

(2) Is PRP cost-effective?—Demonstrating meaningful benefits is insufficient justification for PRP's widespread dissemination; researchers must also demonstrate that the program is a good investment of resources. Findings from a recent study support the cost-effectiveness of a CBT-based depression prevention program similar to PRP (Lynch et al., 2005). This finding is encouraging and should prompt similar studies evaluating PRP's cost-effectiveness. Researchers should consider the cost-effectiveness of PRP in relation to attention-control conditions and alternative interventions.

The cost of PRP's delivery depends on many factors including the method of its delivery. There are benefits and drawbacks to both universal and targeted prevention approaches (see Offord, Kraemer, Kazdin, Jensen, & Harrington, 1998). Targeted interventions, for example, have high costs associated with identifying and recruiting at-risk participants. However, universal prevention requires a greater number of intervention group leaders, which increases compensation and training expenses. Future research should consider the cost-effectiveness of universal and targeted prevention strategies separately. It is important to consider the potentially wide-range of benefits listed in the previous section when evaluating PRP's cost-effectiveness. Small improvements in a variety of domains could translate into large overall benefits, subsequently improving cost-effectiveness estimates.

(3) How does PRP work?—Uncovering the causal mechanisms responsible for PRP's effects on depressive symptoms should be priority for future research. Theoretically, PRP works by improving cognitive style and coping skills. A number of PRP studies have taken steps to test this causal model of change with mixed findings. At least four PRP studies have evaluated cognitive style as a mediator of PRP's effects on depressive symptoms, and three found at least partial support for the model (Cardemil, et al., 2002; Gillham et al., 1995; Roberts, et al., 2004; Yu & Seligman, 2002). Several studies did not find significant intervention effects on depressive symptoms or cognitive style, precluding mediation analyses. Future meta-analyses should attempt to test whether the hypothesized mediation model holds across studies. Researchers seldom report the data needed to test mediation in meta-analysis, making this a difficult task.

The limited data available provide no evidence that PRP is superior to active control conditions that do not target cognitive risk factors. This is consistent with findings from a previous review of depression prevention studies (Merry, et al., 2004). The simplest explanation for this finding is that PRP's effects on depressive symptoms are attributable to factors other than its CBT-based training, like increased attention, expectation of benefit, or group cohesion. It is important, however, to examine the possibility that PRP's effects *are* attributable to CBT skills and that the active comparison conditions produced comparable effects via other mechanisms. Future studies should continue evaluating potential mediators of the effects of both PRP and active control conditions. Doing so will advance our understanding of why these programs produce benefits (when they indeed do produce benefits). PRP researchers should consider whether active control conditions are effective, and less costly, alternative interventions.

There is evidence from dismantling studies in the depression treatment literature that the behavioral, not cognitive, components of CBT may be primarily responsible for treatment gains (Jacobson, et al., 1996). PRP teaches a variety of behavioral coping and problem-solving skills, but few studies have evaluated the cognitive and behavioral program components separately. Future studies should examine these behavioral skills as potential mediators of PRP's effects. If behavioral components are the active ingredient, it may be prudent to revise the program putting greater emphasis on these skills.

(4) Is PRP effective when delivered under real-world conditions?—Intervention effects often decline when programs are transported from university research centers to real-

world delivery settings (Weisz et al, 1995). Our finding that community leaders can deliver PRP effectively is an important step towards effective dissemination. However, this finding alone is not sufficient evidence of PRP's ability to produce effects in community settings. In most studies evaluating community group leaders, the PRP intervention developers provided direct training and ongoing supervision to the group leaders. Such training and supervision may not be feasible if the program is widely disseminated. There were too few studies to evaluate PRP's effectiveness when delivered by community leaders who did not receive direct training and supervision from the program developers. More research is needed to determine the type of training required for leaders to deliver PRP effectively. We encourage PRP researchers to carefully document their group leader training procedures so that future meta-analyses can evaluate whether training methods influence intervention outcomes. Additionally, more research is needed to determine if PRP is effective when implemented under real-world conditions (i.e., in schools and other community settings).

Limitations

This review had several notable limitations. First, we lacked statistical power to evaluate moderators and diagnostic outcomes reliably. Second, we had insufficient data to examine PRP's theoretical model of change (i.e., that improvements in cognitive style and coping skills mediate intervention effects on depression-related outcomes). Third, we had insufficient data to evaluate important outcomes of interest, like adaptive functioning. Finally, we used an ES statistic (d) that assumes normality in the distributions of the two groups under comparison (Acion, et al., 2006). This ES statistic is commonly used in intervention research, including recent meta-analyses of depression prevention programs (e.g., Horowitz & Garber, 2006). Scores on depression measures are rarely normally distributed in non-clinical samples, however; distributions tend to be positively skewed because many participants have few or no symptoms. This may have led to biases in our mean ES estimates. Although there are ES statistics that make less restrictive assumptions, like PS , the information necessary for their calculation (e.g., a U statistic) is rarely reported in intervention studies.

Conclusion

This review confirms that adolescents who participate in the Penn Resiliency Program have fewer depressive symptoms than participants in no-intervention control conditions as late as 12-months post-intervention. While it is encouraging that PRP has enduring effects on symptoms, average effects are small. The top priority of future PRP research should be to determine whether PRP has a meaningful impact on the lives of its participants. Research should examine whether PRP improves adaptive functioning and quality of life, and reduces risk for major mental health problems. PRP aims to provide youth with skills that will help them navigate through adolescence, a time of greatly increased risk, without succumbing to depression and its sequelae. Yet most PRP research has not followed participants past early adolescence. Future research should evaluate PRP's effects throughout the adolescent years. Other priorities include identifying mediators and moderators of PRP's effects and demonstrating that the program is transportable and cost-effective.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge Brad Bushman for providing guidance on conducting meta-analytic analyses; Chris Peterson and Michelle Van Etten Lee for providing helpful comments; Esteban Cardemil, Maranne Wass, and Muriel MacKenzie for graciously providing data from their evaluations of PRP; and Abby Jenkins and Rachel Abenavoli for assisting in data collection. We thank the adolescents, parents, group leaders, administrators, and researchers who have participated in PRP studies over the years. Finally, we are grateful to the National Institute of Mental Health for funding much of the previous research on PRP (through grant # MH52270).

References

References marked with an asterisk provided data used in the meta-analysis.

- Acion L, Peterson JJ, Temple S, Arndt S. Probabilistic index: An intuitive approach to measuring the size of treatment effects. *Statistical Medicine*. 2006; 25:591–602.
- Birleson P. The validity of depressive disorder in childhood and the development of a self-rating scale: A research report. *Journal of Child Psychology and Psychiatry*. 1981; 22:73–88. [PubMed: 7451588]
- Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*. 2006; 26:17–31. [PubMed: 16199119]
- *. Cardemil EV. The prevention of depressive symptoms in inner-city, minority middle school students. (Doctoral dissertation, University of Pennsylvania, 2000). *Dissertation Abstracts International*. 2000; 61:1627.
- *. Cardemil EV, Reivich KJ, Seligman MEP. The prevention of depressive symptoms in low-income minority middle school students. *Prevention & Treatment*. 2002; 5 np.
- *. Cardemil E, Reivich KJ, Beevers CG, Seligman MEP, James J. The prevention of depressive symptoms in low-income, minority children: Two-year follow-up. *Behaviour Research and Therapy*. 2007; 45:313–327. [PubMed: 16643843]
- *. Chaplin TM, Gillham JE, Reivich K, Elkon AGL, Samuels B, Freres DR, et al. Depression prevention for early adolescent girls: A pilot study of all girls versus co-ed groups. *Journal of Early Adolescence*. 2006; 26:110–126. [PubMed: 26139955]
- Comprehensive Meta-Analysis (Version 2) [Computer Software]. Englewood, NJ: Biostat;
- Cooper, H.; Hedges, LV. Potentials and limitations of research synthesis. In: Cooper, H.; Hedges, LV., editors. *The Handbook of Research Synthesis*. New York, NY: Russell Sage Foundation; 1994. p. 521-529.
- Cuijpers P, van Straten A, Smit F, Mihalopoulos C, Beekman A. Preventing the onset of depressive disorders: A meta-analytic review of psychological interventions. *American Journal of Psychiatry*, *AiA*. 2008;1–9.
- Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000; 56:455–463. [PubMed: 10877304]
- *. Gillham JE. Preventing depressive symptoms in school children (Doctoral dissertation, University of Pennsylvania, 1994). *Dissertation Abstracts International*. 1994; 55:4119.
- Gillham JE. Targeted prevention is not enough. *Prevention & Treatment*. 2003; 6 np.
- Gillham, JE.; Brunwasser, SM.; Freres, DR. Preventing depression in early adolescence: The Penn Resiliency Program. In: Abela, JRZ.; Hankin, BL., editors. *Handbook of depression in children and adolescents*. New York, NY: Guilford Press; 2008. p. 309-322.
- *. Gillham JE, Hamilton J, Freres DR, Patton K, Gallop R. Preventing depression among early adolescents in the primary care setting: A randomized controlled study of the Penn Resiliency Program. *Journal of Abnormal Child Psychology*. 2006; 34:203–219. [PubMed: 16741684]

- *. Gillham JE, Reivich KJ. Prevention of depressive symptoms in school children: A research update. *Psychological Science*. 1999; 10:461–462.
- *. Gillham JE, Reivich KJ, Freres DR, Chaplin TM, Shatté AJ, Samuels B, et al. School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn Resiliency Program. *Journal of Consulting and Clinical Psychology*. 2007; 75:9–19. [PubMed: 17295559]
- *. Gillham JE, Reivich KJ, Freres DR, Lascher M, Litzinger S, Shatté A, et al. School-based prevention of depression and anxiety symptoms in early adolescence: A pilot of a parent intervention component. *School Psychology Quarterly*. 2006; 21:323–348.
- Gillham, JE.; Reivich, KJ.; Jaycox, LH. The Penn Resiliency Program. University of Pennsylvania; 2008. Unpublished manuscript
- *. Gillham JE, Reivich KJ, Jaycox LH, Seligman MEP. Prevention of depressive symptoms in schoolchildren: Two-year follow-up. *Psychological Science*. 1995; 6:343–351.
- *. Gillham, JE.; Reivich, KJ.; Seligman, MEP. Prevention of depressive symptoms in children and parents. In: Garber, J., Chair, editor. Preventing depression in youth: Moderators of outcome; Symposium conducted at the annual meeting of the Association for Behavioral and Cognitive Therapies; Philadelphia, PA. 2007.
- Glass, GV.; McGaw, B.; Smith, ML. Meta-analysis in social research. Beverly Hills, CA: Sage; 1981.
- Grissom, RJ.; Kim, JJ. Effect sizes for research: A broad practical approach. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2005.
- Gotlib IH, Lewinsohn PM, Seeley JR. Symptoms versus a diagnosis of depression: Differences in psychosocial functioning. *Journal of Consulting and Clinical Psychology*. 1995; 63:90–100. [PubMed: 7896995]
- Hankin BL. Adolescent depression: Description, causes, and interventions. *Epilepsy & Behavior*. 2006; 8:102–114. [PubMed: 16356779]
- Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*. 1981; 6:107–128.
- Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychological Methods*. 2004; 9:426–445. [PubMed: 15598097]
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003; 327:557–560. [PubMed: 12958120]
- Horowitz JL, Garber J. The prevention of depressive symptoms in children and adolescents: A meta-analytic review. *Journal of Consulting and Clinical Psychology*. 2006; 74:401–415. [PubMed: 16822098]
- Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, et al. A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*. 1996; 64:295–304. [PubMed: 8871414]
- Jané-Llopis E, Hosman C, Jenkins R, Anderson P. Predictors of efficacy in depression prevention programmes: Meta-analysis. *British Journal of Psychiatry*. 2003; 183:384–397. [PubMed: 14594912]
- *. Jaycox LH. Prevention of depressive symptoms and behavioral problems in school children. (Doctoral dissertation, University of Pennsylvania, 1994). *Dissertation Abstracts International*. 1993; 54:6463.
- *. Jaycox LH, Reivich KJ, Gillham JE, Seligman MEP. Prevention of depressive symptoms in school children. *Behaviour Research & Therapy*. 1994; 32:810–816.
- Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry*. 2005; 59:990–996. [PubMed: 16368078]
- Kovacs, M. Children's Depression Inventory manual. North Tonawanda, NY: Multi-Health Systems; 2001.
- Lewinsohn PM, Rohde P, Seeley JR. Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. *Clinical Psychology Review*. 1998; 18:765–794. [PubMed: 9827321]

- Light, RJ.; Singer, JD.; Willett, JB. The visual presentation and interpretation of meta-analyses. In: Cooper, H.; Hedges, LV., editors. *The handbook of research synthesis*. New York, NY: Russell Sage Foundation; 1994.
- Lipsey, MW.; Wilson, DB. *Practical meta-analysis*. Thousand Oaks, CA, US: Sage Publications, Inc; 2001.
- Lynch FL, Hornbrook M, Clarke GN, Perrin N, Polen MR, O'Connor E, et al. Cost-effectiveness of an intervention to prevent depression in at-risk teens. *Archives of General Psychiatry*. 2005; 62:1241–1248. [PubMed: 16275811]
- *. MacKenzie, M.; Kelly, B.; Hunter, SC. Unpublished doctoral thesis. University of Strathclyde; Glasgow: 2008. The promotion of resilience and the prevention of depression: Impact of classroom environment.
- *. Miller JB. The effects of a cognitive-behavioral group intervention on depressive symptoms in an incarcerated adolescent delinquent population. (Doctoral dissertation, The Wright Institute, 1999). *Dissertation Abstracts International*. 1999; 60:1863.
- Merry S, McDowell H, Hetrick S, Bir J, Muller N. Psychological and/or educational interventions for the prevention of depression in children and adolescents. *The Cochrane Library*. 2004; 2 np.
- Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, et al. Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*. 2001; 56:128–165. [PubMed: 11279806]
- Murray CJL, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global burden of disease study. *Lancet*. 1997; 349:1498–1504. [PubMed: 9167458]
- Offord DR, Kraemer HC, Kazdin AE, Jensen PS, Harrington R. Lowering the burden of suffering from child psychiatric disorder: Trade-offs among clinical, targeted, and universal interventions. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1998; 37:686–694. [PubMed: 9666623]
- Overton RC. A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*. 1998; 3:354–379.
- *. Pattison C, Lynd-Stevenson RM. The prevention of depressive symptoms in children: The immediate and long-term outcomes of a school based program. *Behaviour Change*. 2001; 18:92–102.
- Poznanski, EO.; Mokros, H. *Children's Depression Rating Scale Revised (CDRS-R)*. Los Angeles: Western Psychological Services; 1996.
- Prentice DA, Miller DT. When small effects are impressive. *Psychological Bulletin*. 1992; 112:160–164.
- *. Quayle D, Dzuirawiec S, Roberts C, Kane R, Ebsworthy G. The effect of an optimism and lifeskills program on depressive symptoms in preadolescence. *Behaviour Change*. 2001; 18:194–203.
- *. Reivich K. The prevention of depressive symptoms in adolescents (Doctoral dissertation, University of Pennsylvania, 1996). *Dissertations Abstracts International*. 1996; 57:2881.
- Reynolds, WM. *Reynolds Adolescent Depression Scale*. Odessa, FL: Psychological Assessment Resources; 1986.
- *. Roberts C, Kane R, Bishop B, Matthews H, Thomson H. The prevention of depressive symptoms in rural school children: A follow-up study. *International Journal of Mental Health Promotion*. 2004; 6:4–16.
- *. Roberts C, Kane R, Thomson H, Bishop B, Hart B. The prevention of depressive symptoms in rural school children: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*. 2003; 71:622–628. [PubMed: 12795585]
- Roberts RE, Lewinsohn PM, Seeley JR. Screening for adolescent depression: A comparison of depression scales. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1991; 30:58–66. [PubMed: 2005065]
- Rosenthal R. The file drawer problem and tolerance for null results. *Psychological Bulletin*. 1979; 86:638–641.
- Shaffer D, Fisher P, Lucas C, Dulcan M, Schwab-Stone M. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and

reliability of some common diagnoses. *Journal of the Academy of Child & Adolescent Psychiatry*. 2000; 39:28–38.

- *. Shatté AJ. Prevention of depressive symptoms in adolescents: Issues of dissemination and mechanisms of change (Doctoral dissertation, University of Pennsylvania, 1997). *Dissertation Abstracts International*. 1996; 57:7236.
- Spence SH, Shortt AL. Research review: Can we justify the widespread dissemination of universal, school-based intervention for the prevention of depression among children and adolescents? *Journal of Child Psychology & Psychiatry*. 2007; 48:526–542. [PubMed: 17537069]
- Stice E, Shaw H, Bohon C, Marti CN, Rohde P. A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. *Journal of Consulting & Clinical Psychology*. 2009; 77:486–503. [PubMed: 19485590]
- Sutton JM. Prevention of depression in youth: A qualitative review and future suggestions. *Clinical Psychology Review*. 2007; 27:552–571. [PubMed: 17355898]
- *. Tellier JE. Anger and depression among incarcerated juvenile delinquents: A pilot intervention (Doctoral dissertation, The Wright Institute, 1999). *Dissertation Abstracts International*. 1998; 59:4071.
- Wang MC, Bushman BJ. Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods*. 1998; 3:46–54.
- *. Wass M. Evaluation of the Penn Resiliency Program in England. 2008 Unpublished data.
- Weisz JR, Donenberg GR, Han SS, Weiss B. Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*. 1995; 63:688–701. [PubMed: 7593861]
- Woodward, M. *Epidemiology: Study design and data analysis*. 2. Boca Raton, FL: Chapman & Hall/CRC; 2005.
- *. Yu DL. Preventing depressive symptoms in Chinese children. (Doctoral dissertation, University of Pennsylvania, 2000). *Dissertation Abstracts International*. 1999; 60:6389.
- *. Yu DL, Seligman MEP. Preventing depressive symptoms in Chinese children. *Prevention & Treatment*. 2002; 5 np.

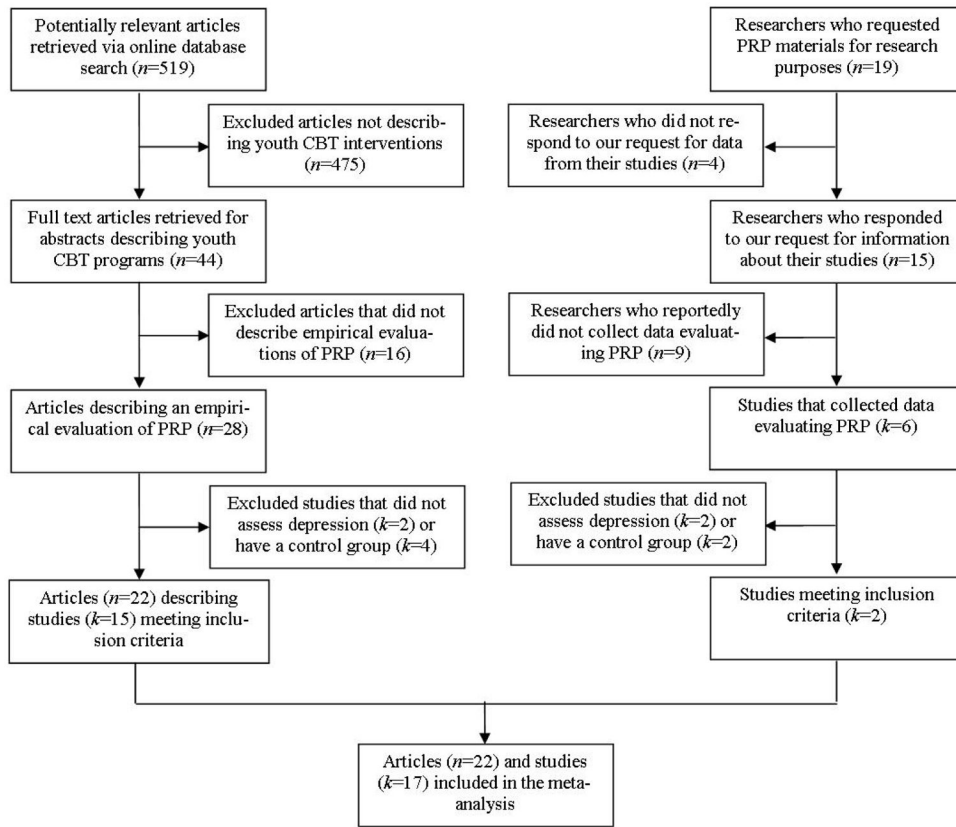


Figure 1. Flow of studies included and excluded from the meta-analytic review.

Table 1
Individual Study Effects (d^U) and Weighted Mean Effects (d_+) on Depressive Symptoms using Random Effects Models.

Study Label	Post-Intervention				6-8-Month Follow-up				12-Month Follow-up			
	n	ω	d^U	95% CI	n	ω	d^U	95% CI	n	ω	d^U	95% CI
Cardemil 2002 Study 1	46	9.9	0.59	0.00, 1.19	41	9.5	0.69	0.06, 1.32	40	9.6	0.61	-0.03, 1.24
Cardemil 2002 Study 2	109	21.6	0.12	-0.25, 0.50	89	21.5	-0.06	-0.48, 0.35	84	21.0	-0.10	-0.53, 0.33
Chaplin 2006	226	36.6	0.29	0.03, 0.55	-	-	-	-	68	17.0	0.08	-0.40, 0.56
Gillham 1994 Study 2	94	18.6	0.21	-0.20, 0.62	50	12.1	0.12	-0.44, 0.68	25	5.8	0.44	-0.37, 1.25
Gillham 2006a	40	9.1	0.08	-0.54, 0.70	35	8.1	0.58	-0.10, 1.26	31	7.5	0.37	-0.34, 1.09
Gillham 2006b	216	35.8	-0.02	-0.29, 0.25	212	48.8	0.22	-0.05, 0.49	193	47.6	0.22	-0.06, 0.50
Gillham 2007a	427	53.4	0.05	-0.14, 0.24	326	72.8	0.06	-0.15, 0.28	327	81.2	0.21	-0.01, 0.43
Gillham 2007b	371	45.6	0.14	-0.08, 0.36	348	73.8	0.12	-0.10, 0.34	322	69.4	0.15	-0.09, 0.39
Jaycox 1994	121	23.4	0.30	-0.06, 0.66	119	29.5	0.30	-0.05, 0.65	85	19.2	0.49	0.04, 0.94
MacKenzie 2008	202	34.2	-0.24	-0.52, 0.04	-	-	-	-	-	-	-	-
Pattison 2001	48	10.2	0.08	-0.50, 0.67	39	9.1	0.49	-0.16, 1.13	-	-	-	-
Quayle 2001	42	9.2	-0.61	-1.23, 0.01	33	7.5	0.60	-0.11, 1.32	-	-	-	-
Reavich 1996	93	18.9	0.04	-0.37, 0.45	93	20.9	0.38	-0.04, 0.81	74	18.3	0.12	-0.34, 0.58
Roberts 2003	179	31.5	0.05	-0.25, 0.34	137	32.5	0.07	-0.26, 0.41	-	-	-	-
Tellier 1998	48	10.4	0.39	-0.19, 0.97	-	-	-	-	-	-	-	-
Wass 2008	21	4.8	0.51	-0.36, 1.38	-	-	-	-	-	-	-	-
Yu 2002	215	35.5	0.23	-0.04, 0.50	207	47.2	0.39	0.11, 0.66	-	-	-	-
Totals	N	$\Sigma\omega$	d_+	95% CI	N	$\Sigma\omega$	d_+	95% CI	N	$\Sigma\omega$	d_+	95% CI
	2498	409.8	0.11	0.01, 0.20	1729	393.3	0.21	0.11, 0.31	1249	296.9	0.20	0.09, 0.32

Note. ω = inverse variance weight; d^U = effect size (unbiased standardized mean difference score); d_+ = mean effect size