2-2018

Identifying Moderators Of Response To The Penn Resiliency Program: A Synthesis Study

S. M. Brunwasser

Jane Gillham

Swarthmore College, jgillha1@swarthmore.edu

HHS Public Access

Author manuscript

Prev Sci. Author manuscript; available in PMC 2017 July 12.

Identifying Moderators of Response to the Penn Resiliency Program: A Synthesis Study

Steven M. Brunwasser and
Vanderbilt University

Jane E. Gillham
Swarthmore College

Abstract

Purpose—To identify moderators of a cognitive-behavioral depression prevention program's effect on depressive symptoms among youth in early adolescence.

Method—Data from three randomized controlled trials of the Penn Resiliency Program (PRP) were aggregated to maximize statistical power and sample diversity ($N = 1145$). Depressive symptoms, measured with the Children's Depression Inventory (CDI; Kovacs, 1992), were assessed at 6 common time points over two-years of follow-up. Latent growth curve models evaluated whether PRP and control conditions differed in the rate of change in CDI and whether youth- and family-level characteristics moderated intervention effects. Model-based recursive partitioning was used as a supplementary analysis for identifying moderators.

Results—There was a three-way interaction of PRP, initial symptom severity, and intervention site on growth in depressive symptoms. There was considerable variability in PRP's effects, with the nature of the interaction between PRP and initial symptom levels differing considerably across sites. PRP reduced depressive symptoms among youth with unmarried parents, but not among those with married parents. Finally, PRP's effects differed across school grade levels.

Discussion—Although initial symptom severity moderated PRP's effect on depressive symptoms, it was not a reliable indicator of how well the intervention performed, limiting its utility as a prescriptive variable. Our primary analyses suggest that PRP's effects are limited to youth whose parents are unmarried. The small number of fifth grade students (n=25; 2%) showed

Correspondence concerning this article should be addressed to Steven Brunwasser, Department of Psychology, Vanderbilt University, 552 Peabody, 230 Appleton Pl, Nashville, TN 37203-5721. steven.brunwasser@vanderbilt.edu.
Steven M. Brunwasser, Vanderbilt University Department of Medicine; Jane E. Gillham, Department of Psychology Swarthmore College and Department of Psychology University of Pennsylvania.

Disclosure of Potential Conflicts of Interest
Dr. Gillham is an author of the Penn Resiliency Program. The authors declare that they have no conflict of interest.

Compliance with Ethical Standards
Ethical Approval
All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.
Informed Consent
Informed consent was obtained from parents of all youth participating in all three randomized controlled trials from which data were drawn for this study. Voluntary assent was obtained from all youth participating in these trials.

a delayed and sustained intervention response. Our findings underscore the importance of evaluating site, family, and contextual characteristics as moderators in future studies.

Keywords

prevention; depression; youth; moderators; cognitive-behavioral

Youth participating in depression prevention programs report lower levels of depressive symptoms and are at reduced risk for depressive episodes compared to peers receiving no intervention. Effects have been inconsistent across trials and even across subgroups within trials (Merry et al., 2011). Research identifying the contexts in which prevention is most (and least) effective could lead to more efficient delivery and inform the development of new intervention strategies for individuals who are unlikely to benefit from existing programs.

A host of moderators have been identified in depression prevention studies. Several studies have found that the magnitude of intervention effects were related to youth characteristics, including initial symptoms levels (Brière, Rohde, Shaw, & Stice, 2014), sociotropy and achievement orientation (Horowitz, Garber, Ciesla, Young, & Mufson, 2007), and gender (Gillham, Hamilton, Freres, Patton, & Gallop, 2006). Other studies have identified family-level characteristics that account for heterogeneity in intervention effects, including parental depression (Garber et al., 2009) and parent-child conflict (Young, Gallop, & Mufson, 2009). There is also evidence that aspects of intervention delivery and contextual factors are linked to potency. For example, higher levels of both intervention dosage and fidelity were predictive of stronger intervention effects in one trial (Gillham et al., 2006). Other studies have found that the magnitude of intervention effects differed across study sites (Beardslee et al., 2013; Gillham et al., 2007). Finally, a meta-analytic review found that prevention effects were larger in trials targeting at-risk youth and in studies with greater proportions of females and youth from ethno-racial minority groups. Additionally, effect size magnitude was positively related to participant age and negatively associated with intervention duration (Stice et al., 2009).

Unfortunately, there is little evidence of consistency in moderators across trials. This is likely attributable in part to challenges that accompany the identification of moderators. Detecting interaction effects requires greater statistical power than the detection of main effects (Shieh, 2009). Power analyses for depression prevention studies tend to be focused on achieving a sufficient sample size for detecting a main effect or an intervention by time effect (Garber et al., 2009); consequently, prevention studies may be insufficiently powered to detect moderators. Additionally, some moderators may operate in complex interactions that are not typically evaluated in linear regression models (Strobl, Malley, & Tutz, 2009). Thus, important intervention moderators may go undetected. On the other hand, moderators detected in individual studies may be specific to the trial's sample and setting, limiting their prescriptive value for future trials. It is critical that researchers employ methods that overcome these challenges.

The Current Study

The primary purpose of this study was to identify participant characteristics and contextual factors that may account for inconsistency in the Penn Resiliency Program's (PRP) effects on depressive symptoms. PRP is a group-based, cognitive-behavioral program for youth in late childhood and early adolescence. There have been at least 20 controlled trials evaluating PRP's effects on depression outcomes. On average, youth participating in PRP report lower levels of depressive symptoms compared to controls receiving no intervention (Brunwasser, Gillham, & Kim, 2009). Although several studies have found main effects on depressive symptoms (Jaycox, Reivich, Gillham, & Seligman, 1994; Yu & Seligman, 2002), others have found effects for only specific subgroups (e.g., Gillham et al., 2007). Some well-powered studies found no effect on depressive symptoms (e.g., Roberts, Kane, Thomson, Bishop, & Hart, 2003).

Multiple PRP trials have found that youth with various risk factors – including elevated baseline depressive symptoms (Gillham et al., 2006), hopelessness (Gillham et al., 2012), and parents with psychopathology (Kindt, Kleinjan, Janssens, & Scholte, 2014) – benefit more from PRP than their peers. One trial found evidence of differential benefit for girls relative to boys (Gillham et al., 2006). In another trial, PRP improved depressive symptoms among adolescents from a predominantly Latino school, and had a fleeting iatrogenic effect among students from a predominantly African American school (Cardemil, Reivich, Beevers, Seligman, & James, 2007). These moderators have not replicated consistently across independent trials.

In this study, we attempted to overcome methodological challenges of identifying intervention moderators. In order to increase sample diversity and statistical power, data from three separate randomized controlled trials (RCTs) of PRP were pooled for analysis. Additionally, in order to address the challenge of selecting and combining candidate moderators, we supplemented our primary analyses with model-based recursive partitioning (MBRP). Recursive portioning-based models allow for complex, non-linear combinations of covariates. Consequently, they may facilitate the identification of moderators that would have gone undetected in linear regression models (Strobl et al., 2009).

Hypotheses

Based on past PRP studies and the broader depression prevention literature, we expected that there would be a positive association between the magnitude of PRP's effect and baseline depressive symptom levels, pessimistic explanatory style, and hopelessness. We also predicted that PRP would be more efficacious for girls and youth in higher academic grade levels (older youth). Finally, we evaluated a number of family-level variables but without a priori predictions.

Method

The three RCTs contributing data in this synthesis study were all considered effectiveness trials because intervention groups were led predominantly or entirely by providers who were naturally present in the intervention settings (e.g., school teachers within schools). In all

three RCTs, no restrictions were placed on outside services for either intervention or control conditions. Detailed descriptions of methods and primary outcomes from each study has been reported elsewhere (Gillham et al., 2006, 2007, 2012). Table 1 provides aggregate demographic information for the three RCTs (see also the online supplement: STables 1–3 and SFigures 1–3).

The first RCT (hereafter, Study 1) evaluated PRP when delivered in primary care clinics by child mental health clinicians (Gillham et al., 2006). Participants (N = 271) were members of a health maintenance organization scoring in the 50th percentile or higher on a depression screening instrument and not meeting criteria for either MDD or dysthymia at screening. Adolescents randomly allocated to PRP attended group sessions at one of two primary care clinics near Sacramento, CA. Those allocated to usual care control received no intervention.

The second RCT (hereafter, Study 2) was conducted in three middle schools in suburban Philadelphia (Gillham et al., 2007). All students who were not actively depressed at baseline were eligible to participate. Participants (N = 697) were randomized to one of three study conditions: no-intervention control, PRP, or the Penn Enhancement Program (PEP). PEP is a placebo intervention designed to mimic non-specific intervention components (Gillham et al., 2007). PRP and PEP groups were led predominantly by teachers and counselors (70% of groups). Graduate students (25%) and research team members (5%) led the remaining groups.

The final RCT (hereafter, Study 3) was also conducted in suburban Philadelphia middle schools (Gillham et al., 2012). Two of the sites participating in Study 2 (sites 4 and 5) also participated in Study 3. Students with elevated depressive symptoms were admitted into the intervention phase of the study first, and others were admitted as space permitted. The final sample (N = 408) had mildly elevated symptoms. Families were allocated randomly to a no-intervention control condition, the standard PRP curriculum, or standard PRP plus a parent intervention program. The vast majority of intervention groups (89%) were led by teachers and counselors; the rest were led by trained research assistants. Given that there were no differential effects between the two PRP conditions (Gillham et al., 2012), they were combined in this study.

In the current study, data from the no-intervention control (CON) and PRP conditions across the three RCTs were aggregated into a synthesized data set (N = 1145). PEP data was excluded because it was only evaluated in Study 2. In terms of sampling and study procedures, studies 2 and 3 were fairly typical of U.S.-based PRP trials. Study 1 was atypical because it was, to our knowledge, the only PRP trial conducted in a primary care setting. The findings from this synthesis study likely provide a less strong representation of PRP's performance in studies conducted outside of the U.S. due to substantial differences in sampling and implementation.

Common Assessments and Measures

The three RCTs shared six common measurement occasions: baseline, immediate post-intervention (post), and follow-ups at 6, 12, 18, and 24 months. Only data from common time points were included. The Children's Depression Inventory (Kovacs, 1992) was the

primary measure of depressive symptoms and was collected at each time point in all studies. Respondents rate the degree to which they have experienced common depressive symptoms on 27 Likert scale items, with higher scores reflecting greater overall symptom severity. In all studies, one item probing for suicidality was removed at the request of school and/or IRB administrators.

Measures of both hopelessness and attributional style were also available in all three RCTs, as measured by the Hopelessness Scale for Children (Kazdin, Rodgers, & Colbus, 1986) and the Composite-Negative (CN) subscale of the Children's Attributional Style Questionnaire (Seligman et al., 1984), respectively. The HSC is comprised of 17 True/False items assessing positive/negative expectations for the future. The CN subscale contains 24 items probing for causal explanations for negative events. In both cases, higher scores indicated greater pessimism.

A dummy variable (*Cond*; 0=CON, 1=PRP) distinguished the two intervention conditions. A number of common variables coding demographic characteristics (hereafter referred to as "common covariates") were available and evaluated as potential moderators of PRP's effect on depressive symptoms. There were three variables describing adolescent characteristics: *Sex* (0=female, 1=male), *Grade* level (0=6th, 1=5th, 2=7th, 3=8th) as a proxy for age, and *Race/*ethnicity (0=Caucasian; 1=African American, 2=Latino/a, 3=Asian, 4=Other). There were four common indicators of parent characteristics: annual family *Income* (0: < $60,000, 1: >= $60,000), maternal and paternal education level (*MomEd* and *DadEd;* 0=no college degree, 1=college degree), and marital status (*Married*; 0= parents unmarried, 1= parents married). For the *Married* variable, "unmarried" included families in which the child's legal parents were divorced, separated, widowed, and never married. Slightly more than half of the participants in the "unmarried" category were divorced or separated (51%). The studies differed in their response options on the martial status questions making it difficult to classify the percentage widowed and never married. Finally, we created indicator variables coding intervention *Study* (Study 1, Study2, or Study 3)*, Site* (identifying to which of the 8 intervention sites participants belonged), and intervention *Setting* (0=primary care clinic, 1=middle school).

Statistical Procedures

Primary analyses—Primary analyses were conducted using Mplus version 7.02 (Muthén, L. K. & Muthén, B. O., 1998). Latent growth curve (LGC) models were used to evaluate whether PRP and CON differed in their rates of growth in depressive symptoms and whether any of the common covariates moderated PRP's effect on growth. Standard errors were adjusted to account for the clustering of participants within studies using a sandwich estimator. Parameters were estimated using robust maximum likelihood (MLR).

In both CON and PRP, depressive symptoms tended to decrease rapidly during the initial follow-ups with the rate of decline decelerating over time. Quadratic growth functions were used to capture the non-constant growth rate. Time was centered at post in all analyses. Thus, the linear slope captured the instantaneous rate of change at post and the quadratic growth slope captured the rate of deceleration across the follow-up. Baseline CDI scores were included as covariates in all models. CDI scores were highly skewed at each

assessment with a substantial percentage of participants with 0 scores (the scale lower limit). Given the skewed distribution and the fact that total scores could take on only positive integer values, CDI total scores were treated as count outcomes with a Poisson distribution. We assessed for overdispersion by comparing the sum of Pearson squared residuals to the residual degrees of freedom (Venables & Ripley, 2002). There were substantial amounts of missing data in the family-level covariates (see Table 1). Multiple imputation (MI) was used to create 10 data sets with no missing values on common covariates (Schafer, 1997). Imputed data sets were based on a quadratic Poisson LGC model in which the growth factors were regressed on all common covariates. Parameter estimates and standard errors were averaged across the imputed data sets (Rubin, 1987).

Recursive partitioning analysis—MBRP models were conducted to supplement our primary analyses. MBRP integrates decision tree methodology into parametric modeling strategies. Traditional decision tree analysis uses recursive partitioning to divide samples into groups with similar values of the dependent variable (DV) using a set of covariates (Strobl et al., 2009). In MBRP, rather than partitioning the sample into groups with similar values of the DV, the sample is partitioned into groups with similar values of a parameter estimate from a parametric statistical model. MBRP identifies covariates (i.e., intervention moderators) that best partition the sample into groups with similar values of the regression coefficient in the parametric model (Strobl et al., 2009; Zeileis, Hothorn, & Hornik, 2008).

In our MBRP analyses we evaluated the degree to which the common covariates moderated PRP's effect on growth in depressive symptoms immediately following the intervention. We first ran a Poisson LGC model with linear and quadratic slopes and saved the factor scores for the growth parameters. The parametric component of the MBRP analysis was linear regression model in which the predicted factor scores for the linear slope of depressive symptoms (*S*; instantaneous change rate) were regressed on *Cond*. The parameter estimates for the regression of *S* on *Cond* were partitioned using all common covariates.

MBRP analyses were conducted using the mobForest package version 1.2 (Garge, Bobashev, & Eggleston, 2013) in R version 3.1.0 (R Core Team, 2014). We used the random forest ensemble method in order to improve model stability. For both MBRP models, we grew forests comprised of 500 regression trees, each based on a random subsample of observations. The portion of the sample randomly excluded in the derivation of each tree served as an "out-of-bag" (OOB) validation sample. In order to maximize diversity in the regression trees in each forest, five covariates were randomly selected as partitioning variables for each tree from the full set of common covariates (Breiman, 2001). Given that there was a substantial amount of missing data on the common covariates, we repeated the MBRP across five imputed data sets (2500 total trees). We limited the MBRP analyses to only 5 rather than 10 imputed data sets because the computational burden for these models was great and model convergence took many hours.

Trees were grown using a conditional permutation method (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008). We assessed the predictive utility of the partitioning variables using the permutation accuracy importance measure (PAIM; Strobl, Boulesteix, Zeileis, & Hothorn, 2007). Larger PAIM values are indicative of greater variable importance.

Partitioning variables were considered potentially important if their value was more than double the absolute value of the partitioning variable with the largest negative value (i.e., the least important moderator), which is a relatively conservative cutoff (Strobl et al., 2009). We report the average mean-squared error ($\overline{MSE}$) and mean pseudo-R^2 (R^2_{pseudo}) for the MBRP model (Breiman, 2001).

Results

Primary Analyses

Table 2 provides descriptive statistics for the CDI at each time point. Exponentiated parameter estimates and confidence intervals are reported below for Poisson growth models to facilitate comprehension of model results. Estimates equal to 1 represent no effect, those less than 1 represent a negative association, and those greater than 1 represent a positive association. The exponentiated estimates are incidence rate ratios (Cameron, 2009): An effect of *Cond* on the intercept of 0.75, for example, indicates that PRP reduced the rate of post CDI points by 25%.

In our primary LGC model (Model 1), we regressed the latent intercept, linear and quadratic slopes (collectively the "growth factors") on all common covariates and all possible two-way interactions between the common covariates and *Cond*. An equation with the full model specification is provided in the online supplement (SEquation 1). The ratio of the sum of Pearson squared residuals to residual degrees of freedom indicated little evidence of overdispersion (ratio = 0.86, $p = .99$). Four moderators emerged as significant in this model. First, CDI^{PRE} moderated PRP's effect on both the linear and quadratic slope factors: $\hat{\gamma}_{112}$ = 1.01, 95% CI [1.0002, 1.02], and $\hat{\gamma}_{212}$= 0.994, 95% CI [0.992, 0.996], respectively. At post, PRP accelerated the rate of instantaneous symptom improvement relative to CON. The Johnson-Neyman method for probing interactions (Preacher, Curran, & Bauer, 2006) showed that PRP's effect on the linear slope was significant across nearly the full range of CDI^{PRE} scores except the top 6%. PRP's effect tended to be larger among youth with low to average symptom levels (see Figure 1a). However, as indicated by the $Cond*CDI^{PRE}$ interaction effect on the quadratic slope, PRP's effect diminished over time, particularly at lower levels of CDI^{PRE}. At 12 months, the peak difference between PRP and CON, the probability that a randomly selected PRP participant would have a better CDI score than a randomly selected CON participant was 56%, 95% CI [52%, 60%] (see Table 1).

Second, there were significant interactions of *Cond*Site* in predicting the growth factors, indicating that the patterns of PRP's effect varied across sites with effects tending to be stronger in sites 1, 2, and 4, and weaker in sites 6 and 8 (see Figure 1b). Sites 1, 2 and 4 were all less affluent than the average (71, 57, and 80% with annual family incomes < $60,000, respectively) and sites 1 and 4 were the most ethno-racially diverse (31 and 52% endorsing an ethno-racial group other than White/Caucasian). Sites 6 and 8 were the most affluent (95 and 88% with annual family incomes >=$60,000, respectively, and > 75% of parents with college degrees). Site 8 was also the most ethno-racially homogenous site (92% White/Caucasian). In an effort to elucidate the *PRP*Site* interactions, we created several pseudo-site-level variables by aggregating over participant-level characteristics (% male, %

married, % of parents with college degree, % ethnoracial minority, mean family income, mean CDI, CN, and HSC scores) and included them as moderators of PRP's effect in an LGC model. We refer to these as pseudo-site-level variables because they represent aggregate characteristics only of the families that chose to participate, but they do not necessarily provide good representations of the sites themselves. There was no evidence that any of the pseudo-site-level characteristics moderated PRP's effect.

Third, the model yielded a significant interaction of *Cond*Married* predicting the linear slope factor ($\hat{\gamma}_{118}$ = 1.58, 95% CI [1.32, 1.89]) such that PRP's effect on the rate of improvement at post was 58% lower (i.e., smaller intervention effect) among youth of married parents relative to youth of unmarried parents. At post, youth of unmarried parents in PRP reported a 52% greater rate of decline in the instantaneous rate of change relative to youth of unmarried parents in CON among youth of unmarried parents ($\hat{\gamma}_{11}$= 0.48, 95% CI [0.36, 0.64]), and they had a 30% lower rate of CDI counts by the 12-month follow-up ($\hat{\gamma}_{01}$= 0.70, 95% CI [0.53, 0.93]). However, there was also a significant effect of *Cond*Married* on the quadratic slope ($\hat{\gamma}_{218}$= 0.80, 95% CI [0.73, 0.87]), indicating that the intervention effect among youth of unmarried parents faded over time (see Figure 2). By the 18-month follow-up there was no longer a significant difference between CON and PRP among youth of unmarried parents ($\hat{\gamma}_{01}$= 0.76, 95% CI [0.56, 1.02]).

Finally, there were significant interactions effects of *Cond*Grade* on both the linear and quadratic slopes ($\hat{\gamma}_{01}$= 2.66, 95% CI [2.05, 3.44] and $\hat{\gamma}_{01}$= 0.50, 95% CI [0.45, 0.54]): At post, the effect of PRP on the instantaneous rate of growth was stronger among 6th graders (the reference group) than 5th graders. However, the intervention effect faded among 6th graders whereas PRP's effect grew stronger and was sustained among 5th graders (see SFigure 4).

A reviewer correctly noted that baseline symptom levels were confounded with both *Study* and *Site* given that studies 1 and 3 prioritized at-risk participants whereas Study 2 enrolled all youth willing to participate. Thus, the two-way interaction of *Cond*CDI*^{PRE} could have been driven by a single study or site. We ran two additional LGC models. In the first, we allowed the effect of PRP on the growth factors to differ across studies (*Cond*CDI*^{PRE}** Study*), but the three-way interaction was not significant. However, in the second model, we allowed PRP's effects on growth to vary across sites (*Cond*CDI*^{PRE}** Site*) and the three-way interaction was significant for several sites relative to the reference level (Site 5), indicating that the strength and direction of the *Cond*CDI*^{PRE} interactions differed across sites. The pattern of intervention effects differed even among sites within the same study. For example, within Study 1, PRP tended to have stronger initial effects at higher levels of *CDI*^{PRE} within Site 1 but the effect magnitude diminished over time. The pattern was very different for Site 2, where effects through the first 12 months tended to be stronger at low to average levels of *CDI*^{PRE} (see STable 3 and SFigure 5).

Model-Based Recursive Partitioning Analyses

The results of the MBRP model were consistent across all five imputed datasets (Figure 3). *CDI*^{PRE} and *Site* were the only partitioning variables to consistently exceed our cutoff for potentially important moderators. Notably, parental marital status and grade level were not

strong predictors in the MBRP model. As a set, the partitioning variables accounted for only a small percentage of the total variability ($\overline{MSE}=0.11$, *SE* = 0.08; $\overline{R}^2_{psudo}=25.72$, *SD* = 0.45).

Discussion

There were three primary findings from this study. First, there was a three-way interaction of condition, pre-intervention symptom levels, and intervention site. Overall, PRP tended to perform better among youth with low to average symptom levels, but the pattern of effects was quite complicated. PRP performed better among symptomatic youth in some sites (e.g., Site 1), and better among low-symptom youth in others (e.g., Site 2), and in some sites, there was no effect regardless of symptom severity (Site 5). Our recursive partitioning analyses confirmed the importance of baseline symptoms and intervention site as moderators of PRP's effect on symptoms. Second, PRP reduced the rate of depressive symptoms among youth whose parents were unmarried, but there was no effect among youth of married parents. Finally, the pattern of PRP"s effect on growth in depressive symptoms differed across grade levels. The effect of PRP among youth in grade 5 was notably weak in the short-term, but the effect grew in magnitude and was better maintained than it was among 6th graders.

Accounting for Differential Response to PRP

Symptom severity and intervention site moderation—That baseline symptom severity and intervention site were moderators of PRP's effect was not surprising. Many depression prevention studies have reported intervention by baseline symptom interactions with prevention effects typically stronger in symptomatic youth (Stice et al., 2009). And intervention site was identified as a moderator in the primary analyses for one of the studies contributing to this synthesis (Study 2; Gillham et al., 2007). Based on findings from past PRP studies and the broader depression prevention literature, we expected that the magnitude of PRP's effect would be positively associated with baseline symptom severity. Our findings, however, suggest that the association between symptom severity and intervention response is far more complex than expected. Consequently, we conclude that baseline symptom severity alone is not a reliable tool for predicting intervention response. Preventionists may choose to prioritize youth with elevated symptom levels with the assumption that intervention effects are more valuable (at least in the short-term) among symptomatic than asymptomatic youth. However, our results suggest that selection of participants based on symptom severity alone may result in uneven findings rather than increased intervention effects.

The most pressing question raised by our findings is: Why do PRP's effects, and the nature of the pre-intervention symptom severity moderation, differ across sites? Unfortunately, our data provide little illumination because site-level characteristics were not coded in these studies. Relative to youth- and family-level characteristics, there has been little exploration of how setting characteristics (e.g., level of school support and enthusiasm for the intervention, recruitment rate, availability of resources and staff) might influence the efficacy of depression prevention programs. This is understandable because most studies

sample participants from only a few intervention settings making it impractical to evaluate whether intervention site characteristics account for variability in intervention response. However, there have been recent PRP trials that have sampled at least a dozen schools (Challen, Machin, & Gillham, 2014; Kindt et al., 2014) that could conceivably provide more information about the role of setting-level characteristics. Given the recent push to make clinical trial data publicly available and advancements in data harmonization (Brown et al., 2013), it would be beneficial for all intervention studies, including those sampling only a small number of intervention sites, to measure site characteristics that could conceivably influence intervention outcome.

It is important to note that group-level characteristics were not included as moderators in this study. Consequently, we do not know whether the heterogeneity in PRP's effects across sites is attributable to site-level characteristics (e.g., availability of resources to support the intervention) or systematic differences in characteristics of the PRP intervention groups across sites (e.g., level of program fidelity or quality of program delivery). It is plausible that the $PRP*CDI^{PRE}*Site$ interaction is largely driven by systematic differences in PRP group characteristics across sites; however, a close inspection of symptom trajectories, particularly among high-symptom participants baseline (CDI > 13), seems to suggest that group characteristics alone are unlikely to explain site inconsistencies. If characteristics of the PRP intervention groups were driving the PRP by site interactions, we would expect to see substantial between-site heterogeneity in growth within the PRP group and larger effects in sites where the PRP groups showed the most improvement. However, symptom trajectories were more consistent across sites for PRP than CON trajectories, particularly among youth with high symptoms (see SFigure 6).

Furthermore, among high-symptom youth, the extent to which the PRP groups showed improvements from baseline through 12 months was unrelated to the magnitude of the 12-month intervention effect ($\hat{\beta}$ = 0.10, 95% CI [−0.44, 0.64]), whereas the mean rate of improvement within CON was a strong predictor of the 12-month intervention effect ($\hat{\beta}$ = −0.50, 95% CI [−0.79, −0.21]). Thus, among high-symptom youth, the amount of symptom improvement within the PRP condition in a given site was a poor indicator of intervention effect magnitude. In most sites, high-symptom youth in CON showed large and sustained reductions in depressive symptoms (regression to the mean). To significantly accelerate the natural recovery process in these sites might not be realistic for a relatively brief group-based intervention, which may explain PRP's generally poor performance among symptomatic youth. However, in sites where the mean symptom trajectory among high-symptom CON participants was worse than the overall mean trajectory across sites (e.g., sites 1 and 4), PRP tended to reduce symptoms. Thus, the best way to ensure that PRP is effective among high symptom youth may be to select participants who are likely to have a more chronic symptom pattern. This could potentially be accomplished by measuring symptoms numerous times prior to participant selection and prioritizing those with sustained elevated symptoms, or by selecting youth with additional risk factors (i.e., indicated +selective prevention). Among low-symptom youth (CDI <= 13), the amount of improvement in both the CON and PRP conditions was predictive of the magnitude of the 12-month effect ($\hat{\beta}$ = 0.95, 95% CI [0.36, 1.54] and $\hat{\beta}$ = −0.96, 95% CI [−1.38, −0.54]).

In summary, baseline symptom severity, in combination with intervention site, was predictive of intervention response. However, without a stronger understanding of the processes driving between-site differences in the nature of the PRP by symptom severity interaction, it is not a very useful predictor of intervention response.

Moderation by marital status—The finding that PRP reduced depressive symptoms only among youth whose parents were unmarried was not anticipated. It may be that youth of unmarried parents are more likely to experience stressors (e.g., family conflict or parental distress) that increase the relevancy of PRP's content. PRP was initially tested in a sample of youth with elevated scores on symptom and family conflict measures (Jaycox et al., 1994). It is noteworthy that two PRP efficacy trials not included in this synthesis study used family conflict in combination with pre-intervention symptom levels in order to select participants at elevated risk for depression. Both studies found robust effects of PRP on both depressive symptoms and explanatory style (Jaycox et al., 1994; Yu & Seligman, 2002). In the Jaycox et al. (1994) study, PRP reduced depressive symptoms among children of divorced parents, though the intervention effect faded by the two-year follow-up (Zubernis, Cassidy, Gillham, Reivich, & Jaycox, 1999).

There are several important caveats. First, we did not have measures of family conflict or parental distress in the aggregated dataset, so we cannot test the hypothesis that children of unmarried parents experienced more family stressors. Second, we used a binary indicator of marital status that did not differentiate between parents who were divorced, separated, widowed, or never married. Youth in all of these groups may experience more family-related stressors than youth of married parents but the nature of these stressors may be different and have different implications for intervention. Additionally, we did not have an indicator of whether parents were married to someone other than the child's other legal parent in the aggregated dataset, but this was the case for only 2% of youth in Study 3. Third, given the large number of moderators tested in our analyses and the fact that marital status was a weak indicator of PRP's effect in the MBRP analyses, we take seriously the possibility that the moderation by marital status was a chance finding. Future PRP studies should measure family distress directly, test whether marital status is reliable marker of family distress, and assess whether there are differential effects of PRP for youth of divorced, separated, and never-married parents.

Moderation by grade level—Depression prevention effects have generally been larger among older adolescents (Stice et al., 2009). Our findings suggest that adolescents in the youngest grade level had a weak initial intervention response; however, they showed a delayed effect that was better sustained over time than those in grade 6. It is plausible that the skills taught in PRP became increasingly relevant for these students during the later follow-ups with the transition to middle school. However, the moderating effect of school grade should be interpreted cautiously for several reasons. First, the sample of youth from grade 5 was very small (n_{CON} = 14; n_{PRP} = 11) and came entirely from Study 1. Second, despite being (presumably) the youngest group in the sample, youth in grade 5 had significantly higher baseline symptoms relative to the full sample, which was unexpected given that depressive symptoms tend to increase with age during adolescence (Hankin et al.,

1998). This suggests that the sample of grade 5 participants was atypical and that findings from this subgroup may not generalize well. Finally, *Grade* was consistently ranked as one of the least prominent moderators in the MBRP analyses. A limitation of this study was that participant age was not available in Study 1 so grade level was used as a proxy in the aggregated data set, decreasing natural variability in age. Within studies 2 and 3, the polychoric correlations between age in years and grade level were .90 and .96, respectively, and the correlation between age in months and grade was 0.84 in Study 2.

Non-Significant Moderators

Our set of partitioning variables accounted for a relatively small portion of the variance in our MBRP model. This suggests that there were unmeasured moderators of PRP's effect. It will be important in future studies to consider alternative moderators (e.g., intervention group and site characteristics) that may account for instability in intervention effects. It is also noteworthy that youth characteristics that had moderated intervention effects in past PRP studies – including trials contributing to this synthesis – did not moderate effects in the aggregate sample despite the fact that there was greater power and multiple methods were used to identify moderators. These moderators appear to have been relevant only in specific studies, limiting their prescriptive value. This study suggests that demographic covariates (e.g., race, sex, family income) and even psychological processes targeted by the intervention (i.e., baseline levels of hopelessness and explanatory style) tell us little about how PRP will perform.

Limitations & Strengths

Limitations of this study include the exclusive reliance on self-report instruments and the availability of only a limited number of potential moderators common to all three RCTs. Notable strengths of this study include the aggregation of data across multiple RCTs and the use of complementary approaches (growth modeling and MBRP) to identify intervention moderators.

Conclusion

Successful implementation of PRP is contingent upon an improved understanding of the conditions and contexts in which the intervention is most successful. With few exceptions, the potential moderators evaluated in this study provided little prescriptive value in terms of selecting participants most likely to benefit from PRP. PRP's performance varies across intervention sites, but we know little about which site characteristics are the most strongly related to intervention response. Priorities for future research include identifying characteristics that will facilitate the selection of youth who are likely to show chronic symptom trajectories and identifying setting characteristics that moderate intervention effects.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Martin Seligman, Karen Reivich, John Hamilton, and Derek Freres for their essential contributions to these projects. We would also like to acknowledge George Howe, Hilda Pantin, Tatiana Perrino, and anonymous reviewers for providing helpful comments on earlier drafts of this manuscript. Finally, above all, we thank the adolescents and families who participated in these research trials.

Funding

This study involved aggregating data from three existing randomized controlled trials. Data from the first trial was obtained from a study funded by the Kaiser Foundation Research Institute. Data from the second and third trials were obtained from studies funded by the National Institute of Mental Health Grant MH52270. Steven Brunwasser was supported in part from an NIMH training grant (T32-MH18921) during completion of this work.

References

Beardslee WR, Brent DA, Weersing VR, Clarke GN, Porta G, Hollon SD, Garber J. Prevention of Depression in At-Risk Adolescents: Longer-term Effects. JAMA Psychiatry. 2013; 70(11):1161–1170. DOI: 10.1001/jamapsychiatry.2013.295 [PubMed: 24005242]

Breiman L. Random forests. Machine Learning. 2001; 45(1):5–32.

Brière FN, Rohde P, Shaw H, Stice E. Moderators of two indicated cognitive-behavioral depression prevention approaches for adolescents in a school-based effectiveness trial. Behaviour Research and Therapy. 2014; 53:55–62. DOI: 10.1016/j.brat.2013.12.005 [PubMed: 24418653]

Brown CH, Sloboda Z, Faggiano F, Teasdale B, Keller F, Burkhart G, Perrino T. Methods for synthesizing findings on moderation effects across multiple randomized trials. Prevention Science. 2013; 14(2):144–156. DOI: 10.1007/s11121-011-0207-8 [PubMed: 21360061]

Brunwasser SM, Gillham JE, Kim ES. A meta-analytic review of the Penn Resiliency Program's effect on depressive symptoms. Journal of Consulting and Clinical Psychology. 2009; 77(6):1042–1054. DOI: 10.1037/a0017671 [PubMed: 19968381]

Cameron, AC. Advances in count data regression: I. Basic cross-section methods. Presented at the Southern California Chapter of the American Statistical Association: 28th Annual Workshop in Applied Statistics; Los Angeles, CA. 2009.

Cardemil EV, 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(2):313–327. [PubMed: 16643843]

Challen AR, Machin SJ, Gillham JE. The UK Resilience Programme: A school-based universal nonrandomized pragmatic controlled trial. Journal of Consulting and Clinical Psychology. 2014; 82(1):75–89. DOI: 10.1037/a0034854 [PubMed: 24245805]

Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TRG, Iyengar S. Prevention of depression in at-risk adolescents: A randomized controlled trial. JAMA: Journal of the American Medical Association. 2009; 301(21):2215–2224. DOI: 10.1001/jama.2009.788 [PubMed: 19491183]

Garge NR, Bobashev G, Eggleston B. Random forest methodology for model-based recursive partitioning: the mobForest package for R. BMC Bioinformatics. 2013; 14(1):125.doi: 10.1186/1471-2105-14-125 [PubMed: 23577585]

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: An Official Publication of the International Society for Research in Child and Adolescent Psychopathology. 2006; 34(2):203–219.

Gillham JE, Reivich KJ, Brunwasser SM, Freres DR, Chajon ND, Kash-MacDonald VM, Seligman MEP. Evaluation of a group cognitive-behavioral depression prevention program for young adolescents: A randomized effectiveness trial. Journal of Clinical Child & Adolescent Psychology. 2012; 41:621–639. DOI: 10.1080/15374416.2012.706517 [PubMed: 22889296]

Gillham JE, Reivich KJ, Freres DR, Chaplin TM, Shatté AJ, Samuels B, Seligman MEP. 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. 2007; 75(1):9–19. DOI: 10.1037/0022-006X.75.1.9 [PubMed: 17295559]

Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. Journal of Abnormal Psychology. 1998; 107(1):128–140. DOI: 10.1037/0021-843X.107.1.128 [PubMed: 9505045]

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(3):401–415. DOI: 10.1037/0022-006X.74.3.401 [PubMed: 16822098]

Horowitz JL, Garber J, Ciesla JA, Young JF, Mufson L. Prevention of depressive symptoms in adolescents: A randomized trial of cognitive-behavioral and interpersonal prevention programs. Journal of Consulting and Clinical Psychology. 2007; 75(5):693–706. DOI: 10.1037/0022-006X.75.5.693 [PubMed: 17907851]

Jaycox LH, Reivich KJ, Gillham JE, Seligman MEP. Prevention of depressive symptoms in school children. Behaviour Research and Therapy. 1994; 32(8):801–816. DOI: 10.1016/0005-7967(94)90160-0 [PubMed: 7993324]

Kazdin AE, Rodgers A, Colbus D. The Hopelessness Scale for Children: Psychometric characteristics and concurrent validity. Journal of Consulting and Clinical Psychology. 1986; 54(2):241–245. DOI: 10.1037/0022-006X.54.2.241 [PubMed: 3700812]

Kindt K, Kleinjan M, Janssens JMAM, Scholte RHJ. Evaluation of a school-based depression prevention program among adolescents from low-income areas: A randomized controlled effectiveness trial. International Journal of Environmental Research and Public Health. 2014; 11(5):5273–5293. [PubMed: 24837666]

Kovacs, M. Children's depression inventory: manual. Multi-Health Systems; 1992.

Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Evidence-Based Child Health: A Cochrane Review Journal. 2011; 7(5):1409–1685. DOI: 10.1002/14651858.CD003380.pub3

Muthén, LK.; Muthén, BO. Mplus User's Guide. 7. Los Angeles, CA: Muthén & Muthén; 1998.

Preacher KJ, Curran PJ, Bauer DJ. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. Journal of Educational and Behavioral Statistics. 2006; 31(4):437–448.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. Retrieved from http://www.R-project.org/

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(3):622–628. [PubMed: 12795585]

Rubin, DB. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley & Sons, Inc; 1987.

Schafer, JL. Analysis of incomplete multivariate data. Chapman&Hall; London: 1997.

Seligman MEP, Kaslow NJ, Alloy LB, Peterson C, Tanenbaum RL, Abramson LY. Attributional style and depressive symptoms among children. Journal of Abnormal Psychology. 1984; 93(2):235–238. DOI: 10.1037/0021-843X.93.2.235 [PubMed: 6725758]

Shieh G. Detecting interaction effects in moderated multiple regression with continuous variables power and sample size considerations. Organizational Research Methods. 2009; 12(3):510–528. DOI: 10.1177/1094428108320370

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 and Clinical Psychology. 2009; 77(3):486–503. [PubMed: 19485590]

Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics. 2008; 9(1):307.doi: 10.1186/1471-2105-9-307 [PubMed: 18620558]

Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics. 2007; 8(1):25.doi: 10.1186/1471-2105-8-25 [PubMed: 17254353]

Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods. 2009; 14(4):323.doi: 10.1037/a0016973 [PubMed: 19968396]

Venables, WN.; Ripley, BD. Modern applied statistics with S. Springer; 2002.

Young JF, Gallop RJ, Mufson L. Mother-child conflict and its moderating effects on depression outcomes in a preventive intervention for adolescent depression. Journal of Clinical Child and Adolescent Psychology. 2009; 38(5):696–704. DOI: 10.1080/15374410903103577 [PubMed: 20183654]

Yu DL, Seligman MEP. Preventing depressive symptoms in Chinese children. Prevention & Treatment. 2002; 5(1)doi: 10.1037/1522-3736.5.1.59a

Zeileis A, Hothorn T, Hornik K. Model-based recursive partitioning. Journal of Computational and Graphical Statistics. 2008; 17(2):492–514. DOI: 10.1198/106186008X319331

Zubernis LS, Cassidy KW, Gillham JE, Reivich KJ, Jaycox LH. Prevention of depressive symptoms in preadolescent children of divorce. Journal of Divorce & Remarriage. 1999; 30(1–2):11–36.

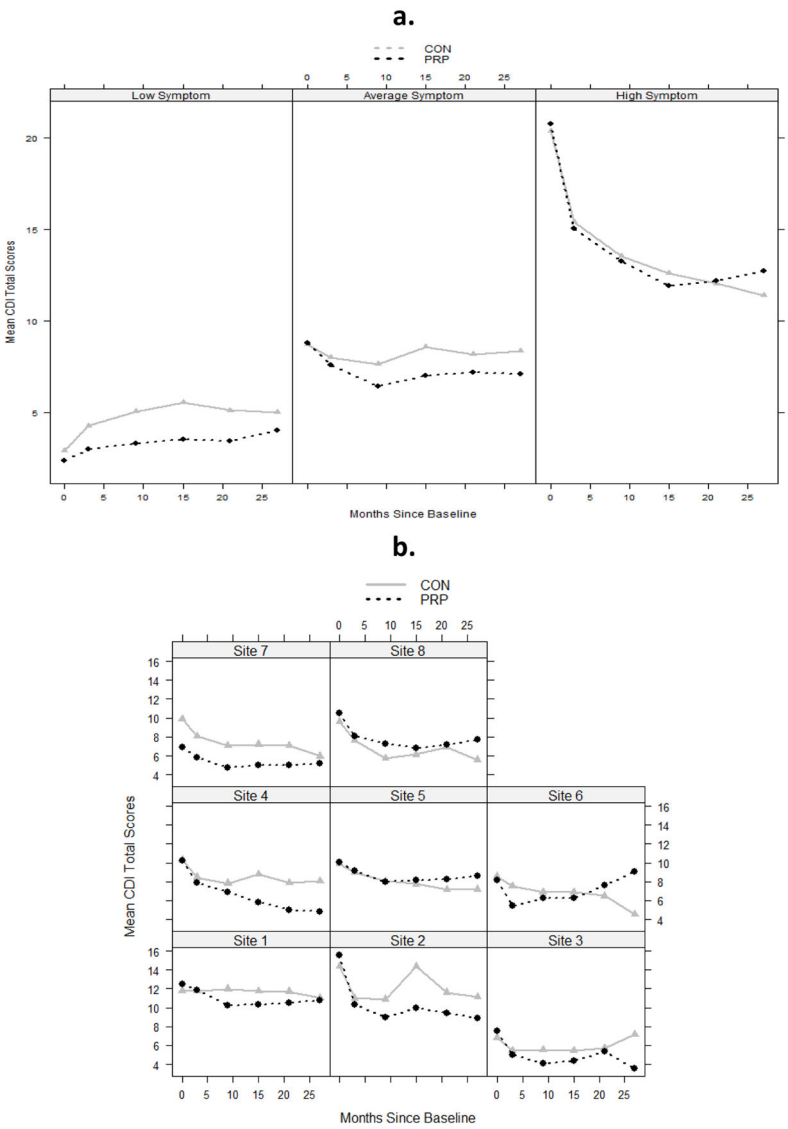

Figure 1. Two-Way Interactions of Cond*CDI^{PRE} and Cond *Site Predicting Growth in Depressive Symptoms

The pattern of PRP's effect varied as a function of both baseline symptom severity (Figre 2a) and intervention site (Figure 2b). In this plot, participants are classified as "Low Symptom" (bottom quartile of *CDI*^{PRE} within the intervention site), "Average Symptom" (25th–75th percentile) or "High Symptom" (top quartile). However, in our analyses baseline *CDI*^{PRE} was a continuous variable. The nature of the interaction *Cond*CDI*^{PRE} interaction also differed across study sites (see SFigures 5 and 6 in the online supplement).

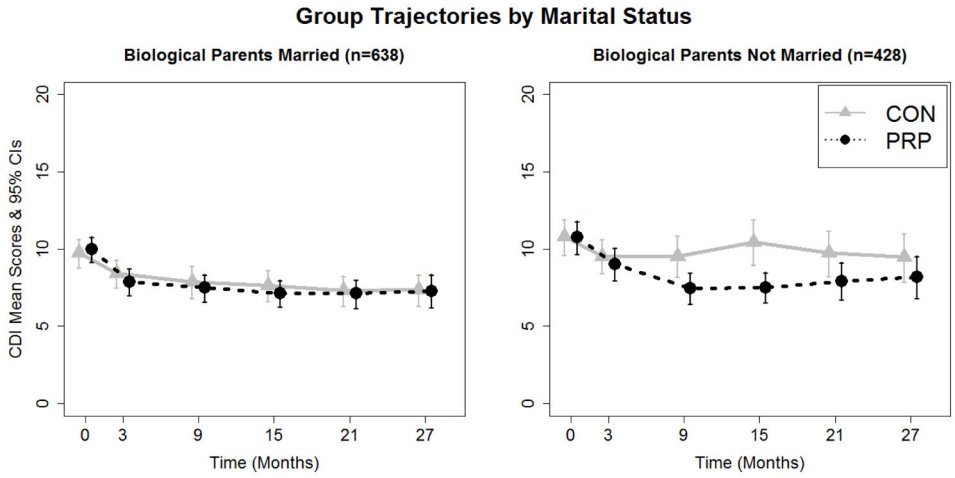

Figure 2. Marital Status Moderates PRP's Effect on Growth in Depressive Symptoms

Among youth whose parents were married at baseline, PRP reduced depressive symptoms relative to CON, but the effect faded and became non-significant by the long-term follow-up assessments. There was no effect among children of married parents.

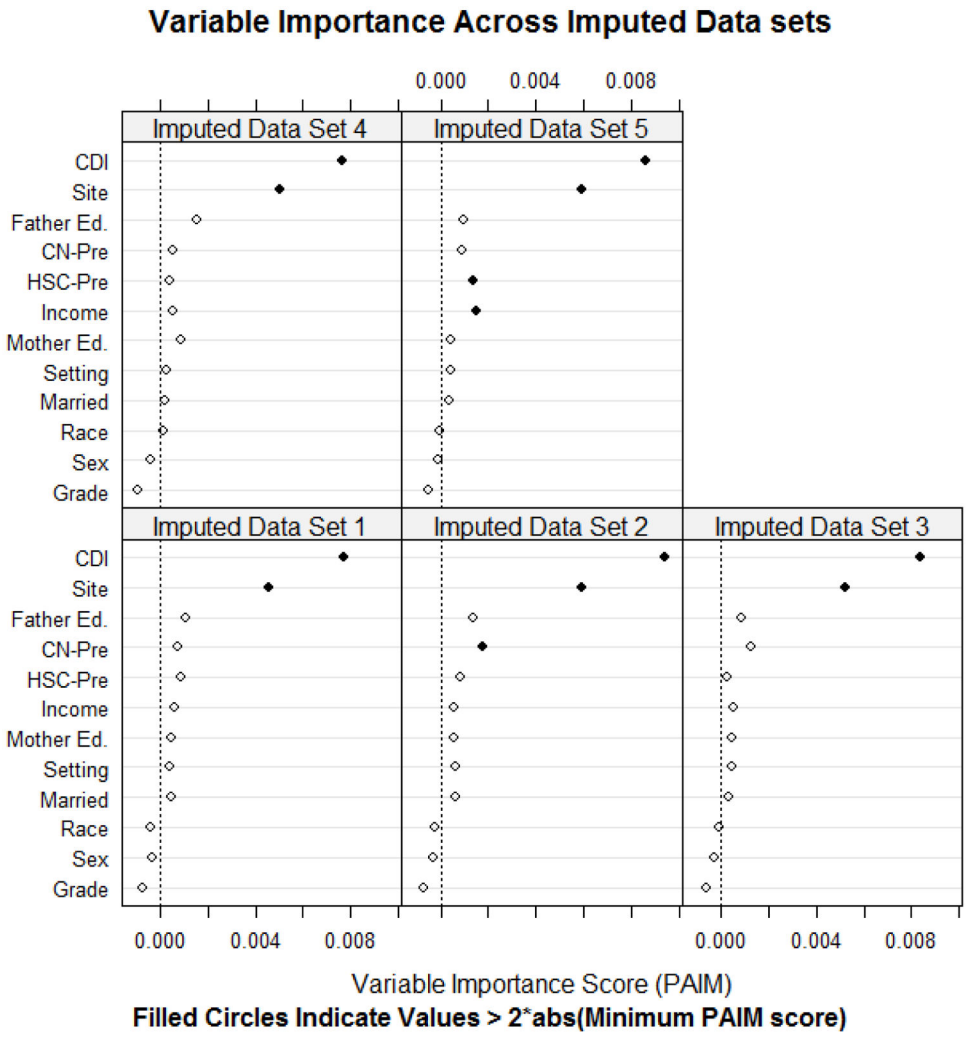

Figure 3. Model-based recursive partitioning (MBRP) importance scores for each imputed data set

Baseline depressive symptoms (CDI) and intervention site were the only consistently strong partitioning variables according to the permutation accuracy importance measure (PAIM).

Table 1

Demographic Characteristics Across Studies on Common Covariates.

	Control (n = 487)		PRP (n = 658)	
	n	**Mean/% (SD)**	**n**	**Mean/% (SD)**
Baseline CDI	487	10.27 (7.87)	657	10.32 (8.14)
Baseline HSC	455	4.71 (2.81)	630	4.37 (2.80)
Baseline CASQ-CN	447	7.77 (3.25)	619	7.59 (3.22)
Grade				
5	14	2.87%	11	1.67%
6	204	41.89%	304	46.20%
7	165	33.88%	210	31.91%
8	99	20.33%	129	19.60%
Unknown/Missing	5	1.03%	4	0.61%
Race				
African				
American/Black	44	9.03%	67	10.18%
Asian/Asian American	19	3.90%	20	3.04%
White/Caucasian	352	72.28%	487	74.01%
Latino/Latina	16	3.29%	28	4.26%
Other	49	10.06%	47	7.14%
Unknown/Missing	7	1.44%	9	1.37%
Father Education				
No College Degree	221	45.38%	317	48.18%
College Degree	154	31.62%	246	37.39%
Unknown/Missing	112	23.00%	95	14.44%
Mother Education				
No College Degree	220	45.17%	289	43.92%
College Degree	172	35.32%	280	42.55%
Unknown/Missing	95	19.51%	89	13.53%
Parent Marital Status				
Married	190	39.01%	238	36.17%
Not Married	268	55.03%	370	56.23%
Unknown/Missing	29	5.95%	50	7.60%
Yearly Family Income				
< $60,000	200	41.07%	272	41.34%
>= $60,000	173	35.52%	278	42.25%
Unknown/Missing	114	23.41%	108	16.41%
Study				
Study 1	124	25.46%	147	22.34%
Study 2	234	48.05%	232	35.26%
Study 3	129	26.49%	279	42.40%
Intervention Setting				

	Control (n = 487)		PRP (n = 658)	
	n	Mean/% (SD)	n	Mean/% (SD)
Middle Schools	363	74.54%	511	77.66%
Primary Care Clinics	124	25.46%	147	22.34%

Table 2

Descriptive Statistics for the Children's Depression Inventory (CDI) Combined Across Trials

	Control			PRP			Comparison
Time	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	***PS* [95% CI][a]**
Baseline	487	10.27	7.87	657	10.32	8.14	.501 [.469, .534]
Post	437	8.95	7.43	577	8.37	7.90	.540 [.506, .573]
6 months	374	8.61	8.00	512	7.48	7.36	.544 [.504, .582]
12 months	358	8.82	8.09	484	7.26	7.13	.557 [.518, .595]
18 months	330	8.35	7.61	437	7.48	7.54	.543 [.501, .584]
24 months	305	8.21	7.63	397	7.64	8.20	.541 [.499, .586]

Note. PRP = Penn Resiliency Program; *PS* = probability of superiority.

[a]Confidence intervals were calculated using bias-corrected bootstrapping.