HHS Public Access

Author manuscript

Eval Health Prof. Author manuscript; available in PMC 2019 June 01.

Published in final edited form as:

Eval Health Prof. 2018 June ; 41(2): 290–320. doi:10.1177/0163278718763499.

Multi-Group Propensity Score Approach to Evaluating an Effectiveness Trial of the New Beginnings Program

Jenn-Yun Tein[1,*], **Gina L. Mazza**[*], **Heather J. Gunn**, **Hanjoe Kim**, **Elizabeth A. Stuart**, **Irwin N. Sandler**, and **Sharlene A. Wolchik**

REACH Institute, Department of Psychology, Arizona State University

Abstract

We used a multi-group propensity score approach to evaluate a randomized effectiveness trial of the New Beginning Program (NBP), an intervention targeting divorced or separated families. Two features of effectiveness trials, high non-attendance rates and inclusion of an active control, make program effects harder to detect. To estimate program effects based on actual intervention participation, we created a synthetic inactive control comprised of non-attenders and assessed the impact of attending the NBP or active control relative to no intervention (inactive control). We estimated propensity scores using generalized boosted models and applied inverse probability of treatment weighting for the comparisons. Relative to the inactive control, NBP strengthened parenting quality as well as reduced child exposure to interparental conflict, parent psychological distress, and child internalizing problems. Some effects were moderated by parent gender, parent ethnicity, or child age. On the other hand, the effects of active versus inactive control were minimal for parenting and in the unexpected direction for child internalizing problems. Findings from the propensity score approach complement and enhance the interpretation of findings from the intention-to-treat approach.

Introduction

Randomized effectiveness trials of parenting interventions often encounter two problems which compromise clear interpretation of findings: low or zero attendance in the intervention and use of active control groups. In many cases and for a variety of reasons, parents who are randomized to a condition will attend either no or only a few sessions. Thus, it is hard to determine whether the program did not work or the program worked but the low exposure was problematic. The use of active control groups can be necessary either because community partners prefer that everyone in the trial receive some credible program, or because there are ethical concerns about not providing a plausibly effective intervention as an alternative to an intervention that has already established efficacy (Leadbeater et al., 2017). In this study, we apply a multi-group propensity score approach to counter the effects of non-attendance and the use of a low-dose active control condition, using data from the

[1]Address Correspondence to: Jenn-Yun Tein, REACH Institute, Department of Psychology, Arizona State University, 900 S. McAllister Ave., Tempe, AZ 85287-6005, Phone: (480) 703-8540, atjyt@asu.edu.

[*]These authors contributed equally to this work

randomized effectiveness trial of the New Beginning Program (NBP) for divorced or separated families.

The study complements the intention-to-treat (ITT) approach, which compared families randomly assigned to the NBP to those randomly assigned to the active control (see Sandler et al., 2017). In randomized control trials (RCTs) with an active comparator, an ITT approach can under- or over-estimate an intervention effect relative to no intervention when either experimental condition suffers from non-attendance (Hernán & Hernández-Díaz, 2012) or when the active control group is either beneficial or harmful. In order to estimate the effects of actually participating in the intervention (rather than simply being randomized to do so), in the current study we compared parents who were randomized to the NBP and attended at least one session and parents who were randomized to the active control and attended at least one session with parents who were randomized to either the NBP or active control but had no exposure to either program because they failed to attend any sessions. This novel approach relies on the creation of a "synthetic inactive control" comprised of non-attenders and the application of a recently-developed multi-group propensity score approach (McCaffrey, Griffin, Almirall, Slaughter, Ramchand, & Burgette, 2013) to evaluate the intervention effects. Supplementing an ITT approach with the propensity score approach provides a more comprehensive assessment of an RCT relative to an ITT approach alone.

The Impact of Divorce

Although the divorce rate has decreased from 1980 (National Center for Health Statistics [NCHS], 2008), parental divorce remains a significant public health concern. Currently, 30–50% of children in the U.S. are expected to experience parental divorce (NCHS, 2008). Numerous studies have shown that parental divorce confers risk for multiple problems in childhood and adolescence, including elevated levels of substance use (Arkes, 2013) and mental health problems (Amato, 2000; Kim, 2011). Furthermore, epidemiologic studies have shown significant relations between parental divorce and substantial increases in clinical levels of substance use and mental health problems, mental health services use, and psychiatric hospitalization in adulthood (e.g., Afifi, Boman, Fleisher, & Sareen, 2009; Hurre, Junkkari, & Aro, 2006).

Several preventive interventions have been shown to reduce substance use and mental health problems in childhood and adolescence when targeting children from divorced families (Pedro-Carroll, Sutton, & Wyman, 1999; Stolberg & Mahler, 1994). The NBP is one of the few evidence-based programs that has been shown to impact multiple domains of functioning in two randomized efficacy trials (Wolchik et al., 1993; 2000; 2002; 2013). Both efficacy trials demonstrated that the NBP significantly improved parenting and reduced child internalizing and externalizing problems at posttest compared to the control condition. The second efficacy trial also showed that these positive effects were evident in the near-term (6 months) and maintained in the long-term (e.g., 6 and 15 years after program completion). Effects were obtained for numerous outcomes including reductions in substance abuse, high risk sexual behavior, and internalizing disorders. The long-term effects of the NBP were accounted for by a cascading mediational model in which immediate effects to strengthen

positive parenting led to improvements in multiple problem domains across developmental periods (Wolchik, Tein, Sandler, & Kim, 2016).

Effectiveness Trial of the New Beginnings Program

Following the two randomized efficacy trials of the NBP, we conducted a randomized effectiveness trial to examine whether favorable program effects could be obtained when the NBP was implemented in community settings. We modified and pilot tested aspects of the NBP to ensure that it would meet the needs of subgroups not included in the efficacy trials, including diverse ethnic groups and fathers. The efficacy trials involved rigorous implementation under ideal circumstances in which the experimenters retained tight control over all aspects of implementation (e.g., resource-intensive settings, homogeneous study population, rigorously trained providers, standardized intervention procedures). However, the effectiveness trial involved collaboration with the family courts to recruit participants, program delivery using community agencies, heterogeneous study population (fathers and mothers, broad age range of children), and fewer eligibility criteria for families (Sandler et al., 2017). For a variety of reasons, each of these natural service delivery factors had the potential to diminish program effects (Schoenwald & Hoagwood, 2001). In addition, two features of the effectiveness trial make it difficult to detect program effects: 1) including an active low-dose control condition and 2) experiencing high levels of non-attendance.

Active control—When conducting effectiveness trials, community-based collaboration and ethical considerations often require a low-dose, treatment-as-usual, or attention placebo control condition in which all participants receive some form of positively-valued intervention (Leadbeater et al., 2017; Popp & Schneider, 2015). In psychosocial research, an active control condition is designed to mimic the theoretically inert elements, but not the active elements, of the intervention (Popp & Schneider, 2015). For example, the active control condition in the NBP effectiveness trial involved discussion of the topics taught in the NBP, building on the parents' own ideas. The facilitator helped parents exchange suggestions for addressing issues that arise post-divorce. Parents did not experience modeling, role-playing, or home practice—features that the program developers considered, but had not tested, as core components of the NBP. Active control conditions are often more effective than no-contact, no-treatment control conditions, such that effects of the NBP versus active control may be smaller in magnitude than effects of the NBP versus no intervention. For example, Chen (2011) reported an overall effect size of Hedges's $g = 0.32$ when comparing attention placebo groups to no-treatment control conditions in a meta-analysis of interventions targeting anxiety or phobia-related problems. Likewise, Merry et al. (2011) found that the effects of preventive interventions for depression were not significant in studies that used an active control but were significant in studies that used an inactive control.

Non-attendance—Non-attendance is a pervasive issue in prevention science, and attendance rates tend to be much lower in effectiveness trials than in efficacy trials (Indrayan & Holt, 2016). Without attending any required sessions, individuals cannot benefit from the intervention. Evaluation of intervention effects with the ITT approach is considered the gold standard for RCTs because it provides the strongest evidence that the effects obtained are

not due to some extraneous factors (other than the intervention itself). However, the ITT approach can confound program efficacy with attendance, particularly in the extreme case where participants randomized to the program never attend a session (Sheiner & Rubin, 1995). Obtaining lower ITT effects than expected brings into question whether the intervention is ineffective or the intervention is effective but diluted by non-attendance (Meier, 1991). Furthermore, Rhew et al. (this issue) point out that failing to attend the intervention is a post-randomization phenomenon, subject to selection and participation bias.

Inactive Control and Propensity Score Approach

The reclassification of parents based on attendance and creation of the synthetic inactive control group nullify the random assignment. The parents who ever attended likely differed from those who never attended, thus confounding group comparisons and threatening the study's internal validity. Propensity scores can be used to adjust for differences in observed baseline characteristics (Rosenbaum & Rubin, 1983). Below, we present some methodological issues associated with the creation of a synthetic inactive control group. We then review propensity scores in general and describe the specific method we used.

Propensity score approach—Randomization ensures that the intervention conditions will not be confounded by either measured or unmeasured baseline characteristics (Shrier, 2013). As a result, intervention effects can be estimated by comparing outcomes directly between intervention conditions. In nonrandomized or observational studies, baseline characteristics often differ systematically across participants of different intervention conditions. One must account for the differences when estimating intervention effects on outcomes; otherwise the estimate may be biased from confounding. Historically, behavioral science researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics across intervention conditions. However, regression adjustment is limited by the number of covariates that can be accommodated and can be biased from misspecifying the relations between the covariates and outcomes (see McCaffrey et al., 2013).

Using a propensity score method to reduce the effects of observed confounding and to examine the causal effects of interventions in nonrandomized or observational studies has become widespread over the past two decades for two-group designs. The propensity score reduces a large set of baseline covariates to a one-number summary. First introduced by Rosenbaum and Rubin (1983), the propensity score is the conditional probability that individual i is assigned to group t based on a set of observed baseline covariates (X_i):

$$e(X_i) = \Pr(Z_i = t \mid X_i)$$

where Z_i represents the intervention group assignment for individual i. Researchers can use $e(X_i)$ to control for imbalances on baseline variables and to reduce confounding effects in nonrandomized studies leading to causal interpretations. Traditionally, logistic regression has been used to estimate propensity scores based on baseline covariates. Newer machine learning techniques (i.e., methods automate analytic model building and use algorithms that iteratively learn from data) such as classification and regression trees (CART; Breiman,

Friedman, Stone, & Olshen, 1984), random forests (Ho, 1998), and generalized boosted modeling (McCaffrey, Ridgeway, & Morral, 2004), have been shown to outperform simple logistic regression models and have become popular alternatives to estimate propensity scores. A detailed review of the methods is beyond the scope of this paper; very tractable explanations can be found in Lee, Lessler, and Stuart (2010); and Strobl, Malley, and Tutz (2009).

The concept of propensity scores originates from the potential outcomes framework (Rubin, 1974). There are two different causal estimands of interest that are relevant for the current focus: one is the "average treatment effect" (ATE), the average causal effect of treatment on the outcome across the full population regardless of whether they received the treatment or not; the other is the "average treatment effect on the treated" (ATT), the average causal effect of treatment on the outcome among the population of those who actually receive the treatment. The choice of using ATE or ATT depends on an investigator's specific research interest (Deb et al., 2016). In our study, we focused on the ATT effects (specifically, the sample average treatment effect on the treated; see Hartman, Grieve, Ramsahai, & Sekhon, 2015)—comparing families who participated in the NBP sessions and families who participated in the active control group sessions, even partially, with families who did not participate in any sessions. Readers can refer to Rubin (1974), Austin (2011), McCaffrey et al., (2013) to understand the fundamental premise of the potential outcomes model, causal estimands, and assumptions of the propensity score approach. The key assumption is known as *unconfounded treatment assignment*, which states that there are no unobserved confounders (factors related to treatment choice and outcomes), once the observed confounders have been adjusted for. This assumption implies that it is crucial to think carefully about the likely confounders and include as many of them as possible in developing the propensity score model.

Once the propensity scores have been estimated, one of four different propensity score methods are commonly used for balancing the distributions of pretest covariates across intervention and control groups and removing the confounding effects: matching, stratification, covariate adjustment, and inverse probability of treatment weighting (IPTW) (see Austin, 2011; Deb et al., 2016; Stuart, 2010). In this study, we applied IPTW to balance the pretest covariates. Analogous to using weights to adjust for sample selection such that individuals in an under-represented group get a higher weight than those in an over-represented group, IPTW creates weights to balance the distributions of pretest covariates between the intervention and control conditions. Weighting each participant by the inverse of his or her propensity score creates a "pseudo sample" in which the distribution of the observed baseline covariates is independent of intervention assignment (Austin, 2011; Deb et al., 2016). Monte Carlo studies conducted by Austin and colleagues (2007, 2011) have shown that matching and IPTW are better at eliminating imbalance between the intervention groups compared to stratification and covariate adjustment. However, matching may not match all of the participants in the control condition, and weighting may be more sensitive to misspecification of the propensity score model (e.g., omitting important confounders such that the *estimated* propensity scores are not adequate realizations of the *true* propensity scores). Furthermore, extreme weight may be placed on participants with a low predicted

probability of receiving the treatment they actually received, leading to unstable or inefficient parameter estimates.

It is critical with IPTW to include the appropriate set of baseline covariates in the propensity score model in order to satisfy the assumption of unconfounded treatment assignment and diagnose balance across the intervention and control groups. Austin and Stuart (2015) stressed that theory and subject matter knowledge should guide the identification baseline covariates related to outcomes (i.e., prognostically important covariates) and related to treatment assignment (i.e., confounders). Several methods have been developed to assess balance in baseline covariates between the intervention and control participants in a sample weighted by the inverse probability of receiving the intervention (see Austin & Stuart, 2015; McCaffrey et al., 2013). If there are imbalances, the model has not been correctly or adequately specified and one can add more covariates or interactions among the covariates to improve the balance.

In the following sections, we briefly describe the NBP effectiveness trial, discuss our propensity score approach with IPTW to achieve covariate balance, and present findings from the program evaluation. The original design addressed three hypotheses: 1) did intervention parents have better quality of parenting and reduced interparental conflict compared to no-intervention parents at immediate posttest and 10-month follow-up? 2) did intervention parents and children report fewer mental health problems than no-intervention parents and children at posttest and 10-month follow-up? 3) were the intervention effects on parenting and child mental health problems moderated by child age, parent gender, parent ethnicity, or baseline status on the outcome? Sandler et al. (2017) presented the results of the ITT analyses comparing the NBP and active control. In the present study, we hypothesized that the intervention elements covered by the active control might produce small changes on parenting or parent and child outcomes. As a result, compared to the inactive control, the NBP and the active control would produce significant effects. We also explored whether the effects varied by the moderators listed above.

Method

Participants

Participants were mothers or fathers from 830 families who were enrolled in the NBP effectiveness trial. Parents were screened based on the following criteria: 1) filing for divorce or separation or, if never married, filing for changes of a parenting time agreement following separation within the past two years; 2) having at least one child aged 3 to 18 with whom the parent spends three or more hours each week or one or more overnights every other week; 3) being able to complete the program and assessments in English; and 4) not being mandated to a parenting class by the Juvenile Court or Child Protective Services. Of the 2,155 parents who expressed interest in the study and who were contacted and screened, 988 (45.8%) met eligibility criteria and 886 (89.7% of those eligible) completed the pretest interview. Data from 56 parents were excluded because their partner was already enrolled in the trial, resulting in a sample of 830 parents (474 [57.1%] mothers and 356 [42.9%] fathers). Parents with multiple 3- to 18-year-old children completed all measures for a randomly-selected “target child” and a subset of measures for all other children. Data were

also obtained from 559 (73.8%) of the 757 eligible children aged 9 to 18 and from teachers of 687 (96.5%) of the 712 children whose parent provided permission for teacher report. In the current study we used only parent report of the target children because small sample sizes in the inactive control for child and teacher report limited our ability to achieve balance across the three conditions.

Study Procedure

Parents were primarily recruited (92.4%) through an invitation given in a brief parenting-after-divorce class, which is mandated for all parents seeking a divorce in Arizona by the family court (ARS § 25–351: "Domestic Relations Education on Children's Issues"). The remaining parents were recruited through media announcements about the program and court referrals. Because the family courts partnered with the study, it was important to offer all parents a program that would be positively received by the parents and would have a reasonable likelihood of being beneficial. Parents were randomized to the NBP (N = 445, 26 mother groups and 24 father groups) or active control condition (N = 385, 22 mother groups and 22 father groups). The intervention protocols were delivered by Master's-level trained facilitators.

NBP—Drawing from social learning and cognitive behavioral theories, the 10-session NBP was designed to promote children's post-divorce adjustment by increasing quality of parenting and decreasing children's exposure to interparental conflict. The program taught skills that had been demonstrated to affect each of these factors (e.g., family fun time for positive parent-child activities, responsive communication, anger management for reducing children's exposure to interparental conflict). Parents role-played the skills in session, practiced the skills at home, and received feedback and support on their home practice.

Active control—In the 2-session active control condition, the same topics taught in the NBP (i.e., challenges of post-divorce parenting, ways to improve parent-child relationship and discipline, ways to reduce children's exposure to interparental conflict) were discussed. Parents set their own goals for changes they would like to accomplish during the program. The facilitator helped parents exchange ideas on how to address issues related to these topics and didactically presented the skills; however, there was no modeling, role-play, or home practice of these skills.

Parents were interviewed before randomization (pretest), immediately following program completion (posttest), and 10 months later (follow-up). Of the 830 parents, 743 (89.6%; 348 active control and 395 NBP) completed the posttest interview and 688 (82.9%; 324 active control and 364 NBP) completed the 10-month follow-up interview. Of the 445 parents in the NBP, 54 (12.1%) attended all 10 sessions, 284 (63.8%) attended between 1 and 9 sessions (M = 6.51, SD = 3.11), and 107 (24.0%) never attended. Of the 385 parents in the active control condition, 265 (68.8%) attended both sessions, 55 (14.3%) attended 1 session, and 65 (16.9%) never attended. We re-categorized these families into three groups: NBP attender group (NBP; parents attended any of the 10 sessions; N = 338), active control attender group (Active Control; parents attended either of the two sessions; N = 320), and

non-attender group (Inactive Control; parents who were assigned to the NBP or active control but never attended; $N = 172$).

Measures

Other than the demographic measures that were assessed only at the pretest assessment, all of the measures discussed below were administered at the pretest, posttest, and 10-month follow-up assessments. The baseline scores on these variables were used as confounders/covariates, and the posttest and 10-month follow-up scores were modeled as outcomes.

Demographics—Parents reported their gender, age, race/ethnicity, highest level of education, legal marital status prior to the divorce or separation, and county of residence. Parents were classified as non-Hispanic White (59.4%), Hispanic (31.4%), or some other race or ethnicity (9.2%). Parents also reported the target child's gender and age.

Parenting quality—We evaluated a broad range of parenting skills and used confirmatory factor analysis (CFA) to model two theorized dimensions of parenting skills: parent-child relationship quality and discipline.

Parent-child relationship quality—We assessed parent-child relationship quality using the 10-item Open Communication scale (Barnes & Olson, 1982; α =.80–.81 across pretest, posttest, and 10-month follow-up), one item assessing parent-child closeness ("How close do you feel to your child?"; Menning, 2006), seven items assessing family routines (Wolchik et al., 2000; α = .77–.81) adapted from the Family Routines Inventory (Jensen, James, Boyce, & Hartnett, 1983), 16-item Acceptance subscale (α = .86–.88) and 16-item Rejection subscale (α = .74–.79) of the parent report version of the Child Report of Parenting Behavior Inventory (CRPBI; Schaefer, 1965), and 9-item Involvement Scale (Menning, 2006; reliability is not applicable for involvement in unrelated activities such as going to a movie, playing a sport).

Discipline—We assessed discipline included the 8-item CRPBI Consistent Discipline subscale (α = .82–.84), 11-item Follow-Through subscale from Oregon Discipline Scale (Oregon Social Learning Center, 1991; α = .77–.80), and a ratio of appropriate to appropriate plus inappropriate use of discipline (appropriate: 9 items, α =.71–.76; inappropriate: 5 items, α =.68–.72) from the Oregon Discipline Scale.

To achieve parsimony and reduce measurement error, we conducted CFA to test a two-factor model of parent-child relationship quality and discipline at each assessment. We allowed involvement and family routines to correlate due to their conceptual similarity and acceptance and consistency of discipline to correlate due to being measured by items on the same scale. The initial model showed that rejection did not fit with the other measures of parent-child relationship quality. The two-factor model excluding rejection adequately fit the data [pretest: $\chi^2(17) = 69.86$, RMSEA = .06, CFI = .96; posttest: $\chi^2(17) = 57.77$, RMSEA = .06, CFI = .97; 10-month follow-up: $\chi^2(17) = 41.25$, RMSEA = .05, CFI = .97]. From this CFA, we used factor scores to create two parenting measures: parent-child relationship quality and discipline. We analyzed rejection separately as a third parenting measure.

Child exposure to interparental conflict—We used four items from the Children's Perception of Interparental Conflict Scale (Grych, Seid, & Fincham, 1992) plus two items that assessed children's exposure to interparental conflict (e.g., "your children saw you and [your ex] yelling, pushing, or shoving each other"; α = .75–.82 for the six items).

Parent psychological distress—We used the 27-item Demoralization subscale from the Psychiatric Epidemiology Research Interview (PERI; Dohrenwend, Shrout, Egri, & Mendelsohn, 1980) to assess parents' general psychiatric symptoms (e.g., "How often have you been bothered by feelings of restlessness?"; α = .93–.93).

Risk—Parents responded to the 15-item Child Risk Index for Divorced or Separated Families (Tein, Sandler, Braver, & Wolchik, 2013; α = .72–.75). The index includes items representing child behavior problems and family-level risk and protective factors (e.g., "Your child has difficulty concentrating," "You and your ex argued about child discipline practices") that are related to child behavior and substance use problems with good predictive validity extending to six years post-assessment (Tein et al., 2013).

Child internalizing, externalizing, and total problems—We assessed children's mental health problems using the internalizing and externalizing subscales of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001; α = .89–.90 for the Externalizing subscale; α = .88–.89 for the Internalizing subscale) for children ages 6 to 18 years old and the parallel subscales of the Preschool Child Behavior Checklist (Pre-CBCL; Achenbach & Rescorla, 2000; α = .91–.92 for the Externalizing subscale; α = .90–.91 for the Internalizing subscale) for children ages 3 to 5. We calculated T scores separately for all CBCL and Pre-CBCL measures based on child age and gender and combined them to assess internalizing, externalizing, and total problems (T. M. Achenbach, personal communication, 2015).

Baseline covariates for generating propensity scores—Following suggestions by Austin and Stuart (2015), we created propensity scores based on 26 baseline covariates theoretically related to 1) both group membership and the outcomes or 2) the outcomes (see Table 1).

Analytic Strategies

Propensity score methods have been primarily applied to studies containing two treatment conditions; however, recently several researchers have extended the methods to three or more conditions (Imai & van Dyk, 2004; Imbens, 2000; McCaffrey et al., 2013). In this study, we generated propensity scores via multiple-treatment generalized boosted models (GBM; McCaffrey, Ridgeway, & Morral, 2004; McCaffrey et al., 2013) and applied IPTW to compare both the NBP and Active Control with the synthesized Inactive Control using the Toolkit for Weighting and Analysis of Nonequivalent Groups (*twang*) package in R (R Development Core Team, 2013; Ridgeway, McCaffrey, Morral, Burgette, & Griffin, 2017). Briefly, *twang* uses GBM to calculate the propensity score weights to estimate pairwise ATT effects. GBM is a machine learning technique, applying an iterative process with multiple regression trees to capture complex relationships between intervention conditions and the baseline covariates. Using several stopping rule criteria, GBM aims to apply the optimal

number of iterations (i.e., number of trees) and estimate weights that yield the best balance of the baseline covariates (McCaffrey et al., 2013).

We used the 26 baseline covariates listed in Table 1 to calculate the propensity score weights. We performed diagnostic checks to examine differences on the baseline covariates across the intervention conditions using two criterion methods (i.e., stopping rules) in *twang*: comparing the mean or maximum of the absolute standardized mean differences between the intervention conditions (i.e., the effect size statistic) as well as the distributions of the covariates between the intervention conditions (i.e., the Kolmogorov-Smirnov statistic). Balance on the baseline covariates after weighting was similar across the two stopping rules. We evaluated whether all pairwise standardized mean differences were ≤ 0.20 as an indication of balance across the three groups (McCaffrey et al., 2013). Baseline covariates with standardized mean differences greater than 0.15 after weighting were included in the outcome models (consistent with the doubly robust method; Kang & Schafer, 2007).

For the missing data on the baseline covariates, we applied mean imputation across groups. The maximum number of missing scores on a baseline covariate was 6 out of 830 (M = 1.25 per baseline covariate). Given these very low rates of missing data, we deemed mean imputation to be an easily implementable and appropriate missing data strategy for the baseline covariates (Cham & West, 2017). We conducted univariate and multivariate outlier analyses to identify influential data points. We also assessed the equivalence of the demographic and pretest variables across the three groups (NBP, Active Control, Inactive Control) before propensity score weighting using one-way analysis of variance for continuous variables and multinomial logistic regression for categorical variables. We also compared standardized mean differences across groups before and after weighting.

We estimated intervention effects using multiple regression, comparing the means of the parent and child outcomes at posttest and 10-month follow-up, separately, weighted by the propensity scores produced by *twang*. Although the programs were delivered in a group-based format, intraclass correlations by intervention group were relatively small (M_{ICC} = .02 across pretest, posttest, and 10-month follow-up; average cluster size = 8.83). We used single-level analyses because the ICCs were small (Stapleton, 2013) and the Inactive Control parents did not interact with other members of their intervention group. We tested the models in M*plus* 7.4 (Muthén & Muthén, 1998–2017) while using full information maximum likelihood estimation to handle missing data on the outcomes. We created two dummy variables denoting intervention conditions, using the Inactive Control as the reference group (i.e., NBP versus Inactive Control and Active Control versus Inactive Control). In each outcome model, we included baseline status on the outcome as a covariate.

In previous trials of the NBP, we found a few significant program x baseline risk interaction effects on child outcomes, with stronger program effects being observed at higher levels of child risk. The NBP was modified and pilot tested to make it appropriate across ethnic groups, parent gender, and age of child. We thus explored whether the intervention effects were moderated by baseline status on the outcome, child age, parent gender, or parent ethnicity (only between non-Hispanic White and Hispanic parents), one moderator at a time. For significant moderation effects, we estimated simple main effects following the

procedures outlined by Aiken and West (1991). Simple main effects are reported at the level where the groups differed significantly for categorical moderators or at one standard deviation below or above (−1 *SD*/+1 *SD*) the mean for continuous moderators. To adjust for conducting multiple tests, we used the false discovery rate (FDR) to control for the expected proportion of false positives among all significant main effects and moderation effects, separately (Benjamini & Yekutieli, 2001). We interpreted effects as reliable if the FDR *p*-value was ≤ .10. We conducted post-hoc power analyses assuming $\alpha = .05$, power $(1 - \beta) = .80$, two-tailed tests of significance, and a correlation of .50 between baseline and post-intervention measures. We had power to detect a small effect size of $R^2 \geq .02$ for each dummy variable or dummy variable by moderator interaction, controlling for covariates.

Results

Tables 2 and 3 reports proportions and means, respectively, for the demographic and pretest covariates used to generate the propensity scores. Before propensity score weighting, significant differences were found for county of residence, parent ethnicity, legal marital status, parent age, parent education, parent-child communication, acceptance, parent binge drinking, and child risk. Compared to parents who attended the NBP or Active Control, parents in the Inactive Control were more likely to be Hispanic, never legally married, younger, and less educated. Parents in the Inactive Control also reported greater parent-child communication and acceptance as well as lower child risk at pretest. No influential data points were identified.

Propensity Score Generation and Balance

We assessed balance on the 26 baseline covariates used to generate the propensity scores by comparing pairwise standardized mean differences across groups before and after weighting. Table 1 shows that all of the standardized mean differences fell below 0.20 after weighting (ranged from 0.002 to 0.452 [median = 0.085] before weighting and from 0.000 to 0.196 [median = 0.045] after weighting. Figure 1 illustrates the maximum pairwise absolute standardized mean differences on the covariates before and after weighting, using a stopping rule based on the Kolmogorov-Smirnov statistic. Standardized mean differences fell between 0.15 and 0.20 after weighting for race/ethnicity, county of residence, parent education, and participation of both parents. These covariates were included in the outcome models.

Program Effects at Posttest

Active control vs. inactive control—Table 4 shows that none of the main effects comparing Active Control and Inactive Control parents at posttest were significant. After the FDR correction, significant effects remained for the following comparisons. Parent gender moderated the effects on child internalizing problems ($B = -6.01$, $SE = 2.12$, $z = -2.84$, $p = .004$), such that Active Control mothers reported significantly higher child internalizing problems ($M_{\text{Active}} = 54.20$, $M_{\text{Inactive}} = 50.18$, $p = .002$) than Inactive Control mothers. None of the effects were moderated by child age, or parent ethnicity.

NBP vs. inactive control—NBP parents reported significantly greater parent-child relationship quality ($B = 0.14$, $SE = 0.06$, $z = 2.24$, $p = .03$) and discipline ($B = 0.28$, $SE =$

0.09, $z = 3.24$, $p = .001$) as well as significantly lower psychological distress ($B = -0.12$, $SE = 0.06$, $z = -2.10$, $p = .04$) and child risk ($B = -0.56$, $SE = 0.26$, $z = -2.14$, $p = .03$) at posttest relative to Inactive Control parents. Parent gender moderated the effects on parent-child relationship quality ($B = 0.26$, $SE = 0.12$, $z = 2.13$, $p = .03$), child exposure to interparental conflict ($B = -0.15$, $SE = 0.07$, $z = -2.25$, $p = .03$), child risk ($B = -1.05$, $SE = 0.51$, $z = -2.07$, $p = .04$), and child internalizing problems ($B = -5.56$, $SE = 2.18$, $z = -2.55$, $p = .01$). Relative to Inactive Control fathers, NBP fathers reported significantly greater parent-child relationship quality ($M_{NBP} = 0.12$, $M_{Inactive} = -0.19$, $p = .002$), lower child exposure to interparental conflict ($M_{NBP} = 1.25$, $M_{Inactive} = 1.35$, $p = .08$), lower child risk ($M_{NBP} = 5.04$, $M_{Inactive} = 6.27$, $p = .001$), and fewer child internalizing problems ($M_{NBP} = 50.17$, $M_{Inactive} = 53.99$, $p = .04$). No significant differences between the NBP and Inactive Control were found for mother reports on these outcomes. None of the effects were moderated by child age, or parent ethnicity.

Program Effects at 10-Month Follow-Up

Active control vs. inactive control—Table 4 shows that none of the main effects comparing Active Control and Inactive Control parents at 10-month follow-up were significant. Child age moderated the effect on child internalizing problems ($B = -0.64$, $SE = 0.24$, $z = -2.70$, $p = .007$) and parent ethnicity moderated the effect of Active Control (versus Inactive Control) on child internalizing and total problems (internalizing: $B = -6.36$, $SE = 2.20$, $z = -2.88$, $p = .004$; total: $B = -4.28$, $SE = 2.01$, $z = -2.13$, $p = .03$). Relative to Inactive Control parents, Active Control parents reported significantly lower child internalizing problems for children 12 years old or older ($M_{Active} = 52.36$, $M_{Inactive} = 55.37$, $p = .03$). Hispanic Active Control parents reported significantly lower child internalizing problems ($M_{Active} = 51.30$, $M_{Inactive} = 54.82$, $p = .04$) relative to Hispanic Inactive Control parents. However, non-Hispanic White Active Control parents reported significantly higher child internalizing and total problems (internalizing: $M_{Active} = 53.29$, $M_{Inactive} = 50.47$, $p = .04$; total: $M_{Active} = 52.48$, $M_{Inactive} = 49.72$, $p = .04$) relative to non-Hispanic White Inactive Control parents. None of the effects were moderated by baseline status on the outcome or parent gender.

NBP vs. inactive control—NBP parents reported significantly lower child exposure to interparental conflict ($B = -0.08$, $SE = 0.04$, $z = -2.13$, $p = .03$) relative to Active Control parents. However, this finding did not meet the FDR correction criterion. Child age and parent ethnicity moderated the effect of NBP (versus Inactive Control) on child internalizing problems (child age: $B = -0.77$, $SE = 0.24$, $z = -3.25$, $p = .001$; parent ethnicity: $B = -4.99$, $SE = 2.35$, $z = -2.13$, $p = .03$). Relative to Inactive Control parents, NBP parents reported significantly lower child internalizing problems for children 9 years old or older ($M_{NBP} = 51.12$; $M_{Inactive} = 53.82$, $p = .01$). Hispanic NBP parents reported significantly lower child internalizing problems ($M_{NBP} = 50.81$; $M_{Inactive} = 54.82$, $p = .03$) relative to Hispanic Inactive Control parents. None of the effects were moderated by baseline status on the outcome or parent gender.

Discussion

This study used a multi-group propensity score approach to evaluate the effects of the New Beginnings Program. The analytic framework contrasted parents who participated in the NBP or active control condition to parents who did not participate in either condition. The study supplements the ITT analyses of the RCT data that contrasted parents randomized to the NBP to parents randomized to the active control condition, regardless of attendance (Sandler et al., 2017). We discuss the study findings based on limitations of the ITT approach, information provided by supplementing the ITT approach with the propensity score approach, implications for program evaluation, and limitations of the study.

Although causal inferences concerning program effects are best supported by an ITT approach comparing outcomes for groups randomly assigned to conditions, one limitation is that the ITT approach tests the effects of assignment to conditions rather than to actual receipt of the experimental intervention. Participants assigned to the experimental condition often either do not receive the program or receive a partial dose so that the ITT approach may underestimate the true effects of actually receiving the experimental condition. In the current study we used a multi-group propensity score approach to estimate the effects of the NBP on those who received any dose (greater than zero). Other authors have evaluated RCTs by supplementing an ITT approach with a propensity score approach to account for different levels of attendance or implementation (see Crowley, Coffman, Feinberg, & Greenberg, 2014; Eisner, Nagin, Ribeaud, & Malti, 2012; Hill, Brooks-Gunn, & Waldfogel, 2003). In these examples, participants randomized to the experimental condition who highly adhered to the protocol were matched with participants randomized to the experimental condition who poorly adhered to the protocol or participants randomized to the control condition based on their conditional probabilities of adherence. The outcomes of compliers in the treatment condition were then compared to the outcomes of matched participants in the control condition who received no or little intervention.

A second limitation of ITT analysis of RCTs is particularly relevant to effectiveness trials conducted in collaboration with community agencies where for ethical or community relations reasons, an active control condition is used rather than a no-treatment or treatment-as-usual condition. Results from an ITT analysis cannot distinguish between the possibility that receipt of both the intervention and active control programs differ from no intervention (in either a positive or negative direction), that receipt of only one of them differs from no intervention, or that receipt of neither of them differs from no intervention. The implications of the findings from the ITT comparison between the intervention and active control groups would differ substantially depending on which of these findings were supported. For example, one explanation for why no difference was found between the active control and intervention groups could be because they were both superior to no intervention. Alternatively, a significant difference between the intervention and active control could be due to an iatrogenic effect of the active control rather than a beneficial effect of the intervention as compared to no intervention.

Causal inferences from the propensity score approach are supported if sufficient baseline covariates are included to rule out unobserved confounding (Imai & van Dyk, 2012). In the

present study, which statistically adjusted for potential confounding, the results of the propensity score approach enhances interpretation of the findings from the ITT approach.

Propensity Score Approach Findings

A high rate of non-attendance in both the NBP and active control conditions provided a means to derive a synthetic inactive control group comprised of parents who were randomly assigned but received no intervention. As a result, we were able to generate propensity scores and apply IPTW to assess the impact of attending the NBP or active control relative to the inactive control. Ultimately, we were able to evaluate the NBP effects when we implemented the intervention in the real-world settings. In addition, we were able to evaluate whether the scaled-down version of the NBP (i.e., with no modeling, role-play, or home practice) could also produce some positive impact. The results of this study indicated that compared to the inactive control, the NBP strengthened quality of post-divorce parenting as well as reduced child exposure to interparental conflict, parent psychological distress, and child internalizing problems. The effects on child internalizing problems were found 10 months following program completion. Some of the effects were moderated by parent gender, parent ethnicity, or child age. On the other hand, compared to the inactive control condition, the effects of the active control on parenting were minimal and on child internalizing problems were sometimes in the opposite direction than expected. In sum, role-play and home practice of the skills taught in the sessions are important for the NBP to be efficacious. A brief didactic presentation of the skills taught in the NBP is not enough to elicit meaningful change in parenting or child mental health problems.

The results of this study enhance the interpretation of the findings from the ITT approach. Comparisons relative to the inactive control enable us to discuss the ITT effects as a possible underestimation of the effects of receipt of the NBP where the active control had a positive effect or as a possible overestimation of the effects of receipt of the NBP where the active control had an iatrogenic effect based on the propensity score approach. The importance of accounting for the effects of active control groups in randomized trials is highlighted by meta-analyses finding that the effects of interventions differ across studies that use an active control group as compared to those that use a no-treatment control group (e.g., Merry et al., 2011).

Applying inverse probability of treatment weighting with the multiple-treatment generalized boosted models enabled us to balance the three groups on baseline characteristics. Modeling and removing the influence of observed baseline confounders renders the study findings to be less biased. However, the way we categorized the three intervention conditions might not be ideal and might not reflect the full potential of the NBP. The NBP included families where parents attended any of the 10 sessions, such that we might underestimate the effects of a full dose of the NBP. Only 12.1% of parents randomized to the NBP attended all 10 sessions (i.e., fully complied with their assigned treatment), such that the program effects might have been diluted even in the propensity score approach due to a high rate of partial compliance. A propensity score approach also has the potential to assess the effects of full compliance with the intervention as compared with no intervention (Eisner et al., 2012).

It should be noted that the newly developed, coarsened exact matching (CEM; Iacus, King, & Prorr, 2009) procedure is a match method that can be also used to control for the potentially confounding influence of baseline covariates and to reduce imbalance between the treated and control groups for multiple-treatment groups. CEM matches data by coarsening continuous baseline covariates to a bin that is chosen by the researcher to maximize the balance. A comparison of using IPTW and CEM would be an interesting project for future work.

Study Limitations

Several limitations of this study need to be acknowledged. Researchers can only effectively implement propensity score weighting if they have sufficient covariate overlap and a large enough sample size for all groups investigated. Smaller samples in any group could seriously under power a study. We attempted to use a propensity score approach to examine program effects based on child report, but because the inactive control group had a small sample size ($N = 81$; data were collected only from children ages 9 to 18), we could not achieve balance. Second, although we were able to balance on a large set of confounders, it is conceivable that other 'unobserved' confounders can explain the group differences in outcomes. In the current study, we did not randomize participants based on program session attendance. However, in a non-experimental design such as ours with post-randomization adjustment, the inclusion of an extensive set of confounders measured and adjusted for represents perhaps the best way to learn about the effects of actually attending the NBP.

References

Achenbach TM, , Rescorla LA. Manual for the ASEBA Preschool Forms & Profiles Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2000

Achenbach TM, , Rescorla LA. Manual for the ASEBA School-Age Forms & Profiles Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001

Aiken LS, , West SG. Multiple regression: Testing and interpreting interactions Newbury Park, CA: Sage; 1991

Afifi TO, Boman J, Fleisher W, Sareen J. The relationship between child abuse, parental divorce, and lifetime mental disorders and suicidality in a nationally representative adult sample. Child Abuse & Neglect. 2009; 33:139–147. DOI: 10.1016/j.chiabu.2008.12.009 [PubMed: 19327835]

Amato PR. The consequences of divorce for adults and children. Journal of Marriage and Family. 2000; 62:1269–1287. DOI: 10.1111/j.1741-3737.2000.01269.x

Arkes J. The temporal effects of parental divorce on youth substance use. Substance Use & Misuse. 2013; 48:290–297. DOI: 10.3109/10826084.2012.755703 [PubMed: 23363082]

Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behavioral Research. 2011; 46:99–424. DOI: 10.1080/00273171.2011.568786

Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. Statistics in Medicine. 2007; 26:734–753. DOI: 10.1002/sim.2580 [PubMed: 16708349]

Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in Medicine. 2015; 34:3661–3679. DOI: 10.1002/sim.6607 [PubMed: 26238958]

Barnes H, , Olson DH. Parent-Adolescent Communication Scale. In: Olson DH, McCubbin HI, Barnes H, Larsen A, Muxen M, , Wilson M, editorsFamily inventories: Inventories used in a national

survey of families across the family life cycle St. Paul, MN: Family Social Science, University of Minnesota; 1982

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics. 2001; 29:1165–1188.

Breiman L, , Friedman J, , Stone CJ, , Olshen RA. Classification and Regression Tree Boca Raton, FL: Taylor & Francis; 1984

Cham H, West SG. Propensity score analysis with missing data. Psychological Methods. 2016; 21:427445.doi: 10.1037/met0000076

Chen M-K. Doctoral dissertation University of Arizona; 2011 Reviews of empirical studies on attention placebo for anxiety or phobia related problems. Retrieved from http://arizona.openrepository.com/arizona/bitstream/10150/203014/1/azu_etd_11863_sip1_m.pdf

Crowley DM, Coffman DL, Feinberg ME, Greenberg MT. Evaluating the impact of implementation factors on family-based prevention programming: Methods for strengthening causal inference. Prevention Science. 2014; 15(2):246–255. DOI: 10.1007/s11121-012-0352-8 [PubMed: 23430578]

Deb S, Austin PC, Tu JV, Ko DT, Mazer CD, Kiss A, Fremes SE. A review of propensity-score methods and their use in cardiovascular research. Canadian Journal of Cardiology. 2016; 32:259–265. DOI: 10.1016/j.cjca.2015.05.015 [PubMed: 26315351]

Domestic Relations Education on Children's Issues Programs. Arizona Revised Statutes § 25–351 et seq.

Dohrenwend BP, Shrout PE, Ergi G, Mendelsohn FS. Nonspecific psychological distress and other dimensions of psychopathology. Archive of General Psychiatry. 1980; 39:1229–1236.

Eisner M, Nagin D, Ribeaud D, Malti T. Effects of a universal parenting program for highly adherent parents: A propensity score matching approach. Prevention Science. 2012; 13:252–266. DOI: 10.1007/s11121-011-0266-x [PubMed: 22232018]

Grych JH, Seid M, Fincham FD. Assessing marital conflict from the child's perspective: The children's perception of interparental conflict scale. Child Development. 1992; 63:558–572. DOI: 10.2307/1131346 [PubMed: 1600822]

Hartman E, Grieve R, Ramsahai R, Sekhon JS. From sample average treatment effect to population average treatment effected on the treated: combining experimental with observational studies to estimate population treatment effects. Journal of the Royal Statistical Society: Series A. 2015; 178:757–778. DOI: 10.1111/rssa.12094

Hernán M, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. Clinical Trials. 2012; 9:48–55. DOI: 10.1177/1740774511420743 [PubMed: 21948059]

Hill JL, Brooks-Gunn J, Waldfogel J. Sustained effects of high participation in an early intervention for low-birth-weight premature infants. Developmental Psychology. 2003; 39:730–744. DOI: 10.1037/0012-1649.39.4.730 [PubMed: 12859126]

Ho TK. The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998; 20:832–844. DOI: 10.1109/34.709601

Hurre T, Junkkari H, Aro H. Long-term psychosocial effects of parental divorce: A follow-up study from adolescence to adulthood. European Archives of Psychiatry and Clinical Neuroscience. 2006; 256:256–263. DOI: 10.1007/s00406-006-0641-y [PubMed: 16502211]

Iacus SM, King G, Porro G. Causal inference without balance checking: coarsened exact matching. Political Analysis. 2012; 20:1–24. DOI: 10.1093/pan/mpr013

Imbens GW. The role of the propensity score in estimating dose-response functions. Biometrika. 2000; 87:706–710. DOI: 10.1093/biomet/87.3.706

Imai K, van Dyk DA. Causal inference with general treatment regimes: Generalizing the propensity score. Journal of the American Statistical Association. 2004; 99:854–866. DOI: 10.1198/016214504000001187

Indrayan A, , Holt MP. Concise Encyclopedia of Biostatistics for Medical Professionals London, UK: Capman & Hall/CRC; 2016

Jensen EW, James SA, Boyce WT, Hartnett SA. The family routines inventory: Development and validation. Social Science & Medicine. 1983; 17:201–211. DOI: 10.1016/0277-9536(83)90117-X [PubMed: 6844952]

Kang JDY, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical Science. 2007; 22(4):523–539. DOI: 10.1214/07-STS227

Kim HS. Consequences of parental divorce for child development. American Sociological Review. 2011; 76:487–511. DOI: 10.1177/0003122411407748

Leadbeater B, Dishion T, Sandler IN, Mauricio A, Bradshaw C, Dodge K, … Smith E. Ethical challenges in promoting the implementation of preventive interventions: Report of the SPR Task Force. 2017 Manuscript submitted for publication.

Lee BK, Lessler J, Stuart E. Improving propensity score weighting using machine learning. Statistics in Medicine. 2010; 29:337–346. DOI: 10.1002/sim.3782 [PubMed: 19960510]

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; 7:1409–1685. DOI: 10.1002/14651858.CD003380.pub3

McCaffrey D, Griffin BA, Almirall D, Slaugher ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Statistics in Medicine. 2013; 32:3388–3414. DOI: 10.1002/sim.5753 [PubMed: 23508673]

McCaffrey D, Ridgeway G, Morral A. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychological Methods. 2004; 9:403–425. DOI: 10.1037/1082-989X.9.4.403 [PubMed: 15598095]

Meier P. Comment on “Compliance as an explanatory variable in clinical trials” by B. Efron and D. Feldman. Journal of the American Statistical Association. 1991; 86:19–22. DOI: 10.2307/2289709

Menning CL. Nonresident fathering and school failure. Journal of Family Issues. 2006; 27:1356–1382. DOI: 10.1177/0192513x06290038

Muthén LK, , Muthén BO. Mplus user’s guide 7. Los Angeles, CA: Muthén & Muthén; 1998–2017

National Center for Health Statistics. Marriage and divorce 2008 Retrieved from http://www.cdc.gov/nchs/fastats/divorce.htm

Oregon Social Learning CenterLIFT Parent Interview Oregon Social Learning Center; Eugene, OR: 1991 Unpublished Manual

Pedro-Carroll JL, Sutton SE, Wyman PA. A two-year follow-up evaluation of a preventive intervention for young children of divorce. School Psychology Review. 1999; 28:467–476.

Popp L, Schneider S. Attention placebo control in randomized controlled trials of psychosocial interventions: Theory and practice. Trials. 2015; 16:150.doi: 10.1186/s13063-015-0679-0 [PubMed: 25872619]

R Development Core TeamR: A language and environment for statistical computing R Foundation for Statistical Computing; Vienna, Austria: 2013 URL http://www.R-project.org

Ridgeway G, , McCaffrey D, , Morral A, , Burgette L, , Griffin BA. Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the twang package 2017 Jul 1. Retrieved from https://cran.r-project.org/web/packages/twang/vignettes/twang.pdf

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.

Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974; 66:688–701. DOI: 10.1037/h0037350

Sandler IN, Wolchik S, Mazza G, Gunn H, Tein JY, Berkel C, … Porter M. Effectiveness of the New Beginnings Program as delivered by natural community settings. 2017 Manuscript in preparation.

Schaefer ES. Children’s reports of parental behavior: An inventory. Child Development. 1965; 36:413–424. DOI: 10.2307/1126465 [PubMed: 14300862]

Schoenwald SK, Hoagwood K. Effectiveness, transportability, and dissemination of interventions: What matters when? Psychiatric Services. 2001; 52:1190–96. DOI: 10.1176/appi.ps.52.9.1190 [PubMed: 11533392]

Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. Clinical Pharmacology & Therapeutics. 1995; 57:6–15. DOI: 10.1016/0009-9236(95)90260-0 [PubMed: 7828382]

Shrier I. Estimating causal effect with randomized controlled trial. Epidemioloty. 2013; 24:779–781. DOI: 10.1097/EDE.0b013e31829f6d21

Stapleton LM. Multilevel structural equation modeling with complex sample data. In: Hancock GR, , Mueller RO, editorsStructural equation modeling: A second course 2. Charlotte, NC: Information Age Publishing; 2013 521562

Stolberg AL, Mahler J. Enhancing treatment gains in a school-based intervention for children of divorce through skill training, parental involvement, and transfer procedures. Journal of Consulting and Clinical Psychology. 1994; 62:147–156. DOI: 10.1037/0022-006X.62.1.147 [PubMed: 8034817]

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; 14:323–348. DOI: 10.1037/a0016973 [PubMed: 19968396]

Stuart EA. Matching methods for causal inference: A review and a look forward. Statistical Science. 2010; 25:1–21. [PubMed: 20871802]

Tein JY, Sandler IN, Braver SL, Wolchik SA. Development of a brief parent report risk index for children following parental divorce. Journal of Family Psychology. 2013; 27:925–936. DOI: 10.1037/a0034571 [PubMed: 24188087]

Wolchik SA, Sandler IN, Millsap RE, Plummer BA, Greene SM, Anderson ER, … Haine RA. Six-year follow-up of a randomized, controlled trial of preventive interventions for children of divorce. Journal of the American Medical Association. 2002; 288:1874–1881. DOI: 10.1001/jama.288.15.1874 [PubMed: 12377086]

Wolchik SA, Sandler IN, Tein JY, Mahrer NE, Millsap RE, Winslow E, … Reed A. Fifteen-year follow-up of a randomized trial of a preventive intervention for divorced families: Effects on mental health and substance use outcomes in young adulthood. Journal of Consulting and Clinical Psychology. 2013; 81:660–673. DOI: 10.1037/a0033235 [PubMed: 23750466]

Wolchik SA, Tein JY, Sandler I, Kim HJ. Developmental cascade models of a parenting-focused program for divorced families on mental health problems and substance use in emerging adulthood. Development and Psychopathology. 2016; 28:869–888. DOI: 10.1017/S0954579416000365 [PubMed: 27427811]

Wolchik SA, West SG, Sandler IN, Tein JY, Coatsworth D, Lengua L, … Griffin WA. An experimental evaluation of theory-based mother and mother–child programs for children of divorce. Journal of Consulting and Clinical Psychology. 2000; 68:843–856. DOI: 10.1037/0022-006X.68.5.843 [PubMed: 11068970]

Wolchik SA, West SG, Westover S, Sandler IN, Martin A, Lustig J, … Fisher J. The Children of Divorce Parenting Intervention: Outcome evaluation of an empirically based program. American Journal of Community Psychology. 1993; 21:293–331. DOI: 10.1007/BF00941505 [PubMed: 8311029]

Yuan KH, Bentler PM. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. Sociological Methodology. 2000; 30:165–200. DOI: 10.1111/0081-1750.00078

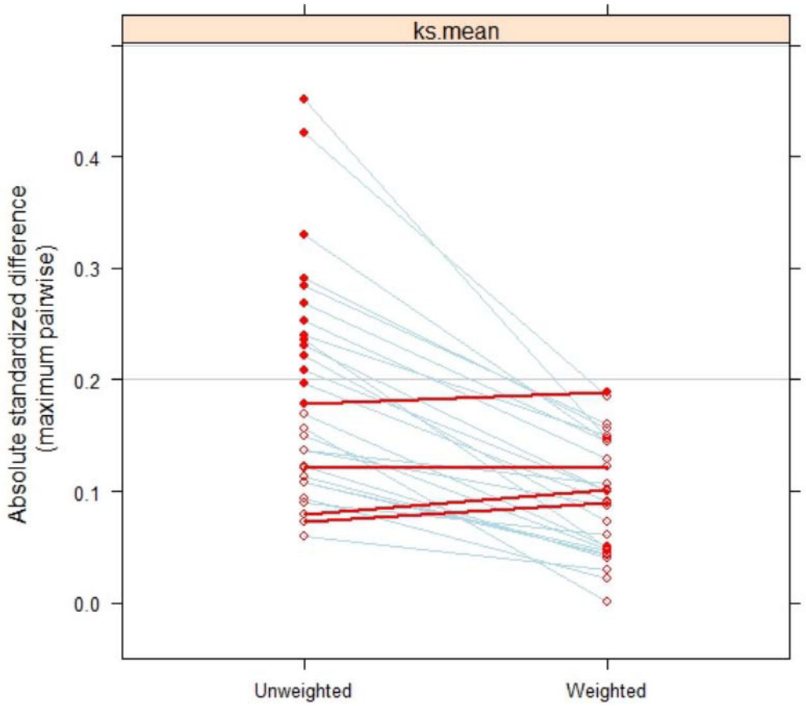

Figure 1.

The maximum pairwise absolute standardized mean differences of the pretest covariates before and after weighting.

Table 1

Standardized Mean Differences Before and After Weighting

Covariate	Before Weighting			After Weighting		
	Inactive vs. Active	Inactive vs. NBP	Active vs. NBP	Inactive vs. Active	Inactive vs. NBP	Active vs. NBP
Participation of Both Parents	0.147	0.179	0.032	0.156	0.196	0.040
County						
Coconino	0.024	0.070	0.094	0.021	0.126	0.105
Yuma	0.045	0.014	0.059	0.076	0.030	0.046
Pima	0.164	0.285	0.121	0.006	0.129	0.123
Race/Ethnicity						
Hispanic	0.290	0.245	0.045	0.152	0.110	0.041
Other Race or Ethnicity	0.015	0.076	0.091	0.040	0.113	0.073
Legal Marital Status	0.215	0.269	0.054	0.069	0.143	0.075
Parent Gender	0.109	0.121	0.013	0.082	0.127	0.045
Parent Age	0.452	0.384	0.069	0.140	0.141	0.001
Child Gender	0.137	0.071	0.066	0.099	0.058	0.041
Child Age	0.156	0.114	0.042	0.010	0.006	0.005
Parent Education	0.327	0.421	0.095	0.136	0.185	0.049
Contact with Child	0.150	0.076	0.074	0.051	0.007	0.043
Parental Involvement	0.071	0.066	0.137	0.111	0.022	0.089
Parent-Child Closeness	0.076	0.093	0.017	0.016	0.018	0.001
Parent-Child Communication	0.223	0.195	0.027	0.083	0.060	0.023
Family Routines	0.085	0.108	0.023	0.048	0.023	0.026
Acceptance	0.202	0.237	0.034	0.027	0.045	0.018
Rejection	0.102	0.113	0.010	0.050	0.035	0.015
Consistency of Discipline	0.053	0.026	0.079	0.089	0.053	0.036
Appropriate Use of Discipline	0.156	0.197	0.041	0.058	0.078	0.020
Follow-Through	0.010	0.076	0.066	0.071	0.041	0.030
Interparental Conflict	0.060	0.049	0.011	0.013	0.000	0.013
Demoralization	0.072	0.004	0.068	0.091	0.056	0.035
Parent Binge Drinking	0.207	0.209	0.002	0.067	0.092	0.025

Eval Health Prof. Author manuscript; available in PMC 2019 June 01.

Covariate	Before Weighting			After Weighting		
	Inactive vs. Active	Inactive vs. NBP	Active vs. NBP	Inactive vs. Active	Inactive vs. NBP	Active vs. NBP
Child Risk Index	0.222	0.231	0.009	0.079	0.098	0.018
Internalizing Problems	0.149	0.169	0.020	0.012	0.052	0.040
Externalizing Problems	0.108	0.106	0.003	0.026	0.045	0.019
Total Problems	0.116	0.122	0.007	0.016	0.042	0.026

Table 2

Comparison of Categorical Demographic Variables Across Groups

Variable	Synthetic Inactive Control	Active Control	NBP	*p*-value
	$\hat{p}$	$\hat{p}$	$\hat{p}$	
Participation of Both Parents				.162
Both Parents	.035	.072	.080	
One Parent	.965	.928	.920	
County				.019
Coconino	.087	.094	.068	
Yuma	.093	.106	.089	
Pima	.448	.369	.311	
Maricopa	.372	.431	.533	
Parent Race/Ethnicity				.016
Non-Hispanic White	.488	.619	.624	
Hispanic	.413	.278	.299	
Other	.099	.103	.077	
Legal Marital Status				.015
Ever Legally Married	.779	.856	.876	
Never Legally Married	.221	.144	.124	
Parent Gender				.400
Male	.384	.438	.444	
Female	.616	.563	.556	
Child Gender				.340
Male	.494	.563	.530	

Table 3

Comparison of Continuous Demographic and Baseline Variables Across Groups

Variable	Synthetic Inactive Control	Active Control	NBP	*p*-value
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	
Parent Age	34.77(8.70)	38.45(8.05)	37.90(7.69)	< .001
Child Age	7.98(3.99)	8.63(4.25)	8.46(4.13)	.252
Parent Education	13.81(2.20)	14.58(2.39)	14.80(2.32)	< .001
Contact with Child	21.38(8.94)	20.08(8.52)	20.72(8.59)	.270
Parental Involvement	4.80(1.01)	4.87(0.93)	4.74(1.08)	.212
Parent-Child Closeness	4.63(0.75)	4.57(0.74)	4.56(0.75)	.597
Parent-Adolescent Communication	4.57(0.48)	4.46(0.52)	4.48(0.48)	.048
Family Routines	2.66(0.38)	2.62(0.35)	2.62(0.37)	.506
Acceptance	4.57(0.48)	4.48(0.45)	4.46(0.48)	.034
Rejection	1.49(0.36)	1.53(0.47)	1.53(0.35)	.451
Consistency of Discipline	4.44(0.52)	4.47(0.49)	4.43(0.56)	.593
Appropriate Use of Discipline	0.46(0.05)	0.47(0.05)	0.47(0.06)	.103
Follow-Through	3.97(0.70)	3.96(0.69)	3.91(0.72)	.488
Interparental Conflict	1.50(0.47)	1.53(0.47)	1.52(0.47)	.810
Demoralization	2.12(0.65)	2.08(0.59)	2.12(0.57)	.624
Parent Binge Drinking	0.58(0.94)	0.41(0.80)	0.41(0.71)	.052
Child Risk Index	6.16(3.14)	6.87(3.30)	6.89(3.12)	.030
Internalizing Problems	54.34(10.88)	55.93(10.73)	56.15(10.64)	.173
Externalizing Problems	53.41(10.32)	54.50(9.93)	54.47(9.92)	.458
Total Problems	54.19 (11.31)	55.42(10.48)	55.49(10.44)	.380

Table 4

Results for Parent Reported Outcomes at Posttest and 10-Month Follow-Up

Variable	Adjusted Mean (NBP)	Adjusted Mean (Active Control)	Adjusted Mean (Inactive Control)	Main Effect [95% CI] Active vs. Inactive/NBP vs. Inactive	Main Effect *p*-value Active vs. Inactive/NBP vs. Inactive	Significant Moderators (*p*-value) Active vs. Inactive/NBP vs. Inactive
Posttest						
Parent-Child Relationship Quality	0.104	−0.092	−0.039	−0.05 [−0.19, 0.08] 0.14 [0.02, 0.27]	.430/**.025***	Parent Gender .097/**.033***
Discipline	0.158	−0.050	−0.118	0.07 [−0.09, 0.22] 0.28 [0.11, 0.44]	.398/**.001***	
Rejection	1.411	1.455	1.466	−0.01 [−0.08, 0.05] −0.06 [−0.12, 0.01]	.724/.089	Baseline **.042/.035**
Interparental Conflict	1.274	1.331	1.271	0.06 [−0.01, 0.07] 0.003 [−0.07, 0.07]	.102/.936	Parent Gender .142/**.025***
Parent Distress	1.869	1.932	1.990	−0.06 [−0.01, 0.07] −0.12 [−0.07, 0.07]	.302/**.036***	Parent Gender **.027**/.141
Child Risk Index	5.232	5.695	5.790	−0.10 [−0.62, 0.43] −0.56 [−1.07, −0.05]	.721/**.032***	Parent Gender .067/**.039***
Internalizing	51.287	53.342	51.534	1.81 [−0.33, 3.95] −0.25 [−2.42, 1.93]	.098/.824	Parent Gender **.004***/**.011***
Externalizing	50.682	52.101	52.146	−0.04 [−1.92, 1.83] −1.46 [−3.35, 0.42]	.963/.128	
Total Problems	50.889	52.915	51.855	1.06 [−0.88, 3.00] −0.97 [−2.91, 0.97]	.284/.329	
10-Month Follow-Up						
Parent-Child Relationship Quality	0.075	−0.015	0.017	−0.03 [−0.19, 0.13] 0.06 [−0.10, 0.22]	.701/.483	
Discipline	0.027	−0.079	−0.064	−0.02 [−0.20, 0.17] 0.09 [−0.10, 0.28]	.875/.337	
Rejection	1.427	1.444	1.458	−0.01 [−0.08, 0.06] −0.03 [−0.10, 0.04]	.707/.393	
Interparental Conflict	1.229	1.249	1.313	−0.06 [−0.01, 0.07] −0.08 [−0.07, 0.07]	.096/**.033**	
Parent Distress	1.821	1.862	1.892	−0.03 [−0.16, 0.10] −0.07 [−0.21, 0.06]	.649/.297	
Child Risk Index	5.216	5.222	5.513	−0.29 [−0.89, 0.31]	.340/.348	

Variable	Adjusted Mean (NBP)	Adjusted Mean (Active Control)	Adjusted Mean (Inactive Control)	Main Effect [95% CI] Active vs. Inactive/NBP vs. Inactive	Main Effect *p*-value Active vs. Inactive/NBP vs. Inactive	Significant Moderators (*p*-value) Active vs. Inactive/NBP vs. Inactive
				−0.30 [−0.92, 0.32]		
Internalizing Problems	51.486	52.760	53.176	−0.42 [−2.56, 1.73] −1.69 [−3.92, 0.54]	.704/.137	Baseline **.041**/.184 Parent Ethnicity **.004***/**.033*** Child Age **.007***/**.001***
Externalizing Problems	50.847	51.024	49.733	1.29 [−0.69, 3.28] 1.11 [−0.94, 3.17]	.202/.288	
Total Problems	51.643	52.112	51.808	0.30 [−1.62, 2.22] −0.17 [−2.12, 1.79]	.757/.868	Parent Ethnicity **.033***/.140

*Note*The asterisk (*) indicates that the false discovery rate *p*-value was ≤ .10.