AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology


ISSN: 0022-006X

2026, Vol. 94, No. 1, 49–62
https://doi.org/10.1037/ccp0000991

# Who Benefits Most From Couple Relationship Education: A Machine Learning Approach

Po-Heng Chen and Hannah C. Williamson
Department of Human Development and Family Sciences, The University of Texas at Austin

***Objective:*** Couple relationship education (CRE) seeks to enhance relationship functioning and prevent deterioration of relationship quality over time. However, impacts of CRE are mixed and often appear to be influenced by the characteristics of the couples receiving the intervention. To provide effective interventions, a better understanding of the couples who are most likely to benefit from CRE is needed. Unfortunately, the existing literature has failed to account for the complex and interdependent nature of pretreatment risk factors, leading to inconsistent and inconclusive results. ***Method:*** The present study addresses this issue by applying causal forest, a machine learning technique, to two randomized controlled trials of CRE to determine the pretreatment characteristics that are most predictive of treatment outcomes. In Study 1, data from 6,298 couples were used to train causal forest algorithms, and in Study 2, data from 1,595 couples were used to test the accuracy and generalizability of the trained models. ***Results:*** Causal forest models indicated that pretreatment characteristics predicted 12-month treatment effects, such that participants with higher psychological distress and lower baseline relationship happiness experienced greater improvements in relationship happiness, while those with higher psychological distress and perceived stress had greater reductions in negative emotions and behaviors within the relationship. These results were robust when tested in a novel data set. ***Conclusions:*** This research highlights the underlying heterogeneity in CRE treatment effects and demonstrates the ability of machine learning methods to identify who may benefit most from CRE and can inform efforts to improve targeting of these interventions.

***What is the public health significance of this article?***

Over the past 2 decades, the federal government has invested more than $2 billion in couple relationship education programs targeting lower income couples in the United States. Identifying the couples who are most likely to benefit from these intervention efforts, as well as those who reap little benefit, is crucial for ensuring that these funds are spent wisely, but currently little is known about how to identify couples who are a good match for this type of intervention. Using data from two randomized trials, this study shows that low-income couples who were experiencing high levels of relational and psychological distress and high levels of external stress were the most likely to benefit from couple relationship education.

*Keywords:* causal forest, close relationships, couples, intervention, relationship education

*Supplemental materials:* https://doi.org/10.1037/ccp0000991.supp



Jessica Hamblen served as action editor.
Po-Heng Chen https://orcid.org/0000-0001-6697-1668
Hannah C. Williamson https://orcid.org/0000-0002-4816-3621
The data reported in this article were obtained from publicly available data archived at the Inter-university Consortium for Political and Social Research. The Supporting Healthy Marriage project is available at https://www.icpsr.umich.edu/web/DSDR/studies/34420. A bibliography of journal articles, working articles, conference presentations, and dissertations using the Supporting Healthy Marriage project is also available through the Inter-university Consortium for Political and Social Research. The Parents and Children Together data are available at https://www.icpsr.umich.edu/web/ICPSR/studies/37843. A bibliography of journal articles, working articles, conference presentations, and dissertations using the Parents and Children Together data is also available through the Inter-university Consortium for Political and Social Research. The variables and relationships examined in the present article have not been examined in any previous or current articles or, to the best of the authors' knowledge, in any articles that will be under review soon.

Analysis code is available at https://osf.io/42qyz/. The studies reported in this article were not preregistered. Portions of this research were presented at the annual convention of the Association for Behavioral and Cognitive Therapies in November 2024 and of the Society for Personality and Social Psychology in February 2025. The authors have no known conflicts of interest to disclose.

Po-Heng Chen played a lead role in data curation, formal analysis, methodology, project administration, software, visualization, and writing–original draft and an equal role in conceptualization. Hannah C. Williamson played a lead role in supervision, validation, and writing–review and editing, a supporting role in formal analysis, methodology, and writing–original draft, and an equal role in conceptualization.

Correspondence concerning this article should be addressed to Hannah C. Williamson, Department of Human Development and Family Sciences, The University of Texas at Austin, 108 East Dean Keeton Street, Austin, TX 78712, United States. Email: hwilliamson@utexas.edu

Intimate relationships are one of the most consequential experiences in our lives (Fletcher et al., 2015). Being part of a healthy intimate relationship conveys a wealth of benefits, including better health, psychological well-being, and economic benefits for the romantic partners, as well as creating a positive environment for children to grow and flourish (e.g., Braithwaite & Holt-Lunstad, 2017; Don et al., 2025; Robles et al., 2014). Despite the far-reaching effects of strong intimate relationships on individuals' lives, maintaining a strong relationship is difficult for many people: 40% to 50% of first marriages in the United States end in divorce, and dissolution is even higher for couples in cohabiting nonmarital relationships (Cherlin, 2010). In recognition of the struggle that many couples face in maintaining healthy, satisfying relationships, scholars developed couple relationship education (CRE) as a preventive intervention that could be widely disseminated in order to prevent the development of relationship distress (Halford et al., 2008).

However, CRE has not been wholly successful in achieving this goal. Estimates of treatment effects for CRE have varied widely, with some studies documenting effect sizes that are moderate in magnitude (e.g., Doss et al., 2016), others finding very small or null results (e.g., Halford et al., 2017), and some even documenting iatrogenic effects (e.g., Rogge et al., 2013). Given the range of potential outcomes, understanding the factors that predict treatment response has long been of interest to the field. However, the existing literature on this topic suffers from methodological limitations, such as small sample sizes, focus on a single moderator, and/or use of parametric analyses with strict assumptions that are likely violated. Therefore, results of this literature have been mixed, leaving us with no clear picture of which couples will benefit most from CRE. The present study addresses this issue by using a cutting-edge machine learning technique (causal forest; Wager & Athey, 2018) that can overcome these methodological limitations and provide a clearer picture of the pretreatment characteristics that are most predictive of treatment outcomes in CRE using data from two large randomized controlled trials (RCTs).

## Moderators of CRE Effectiveness

Some research has already been done to examine whether pretreatment couple characteristics moderate treatment effects, though this literature has largely focused on a single characteristic: relationship distress. Even within this single domain, results are mixed as to which couples benefit the most from CRE. Much of the existing research supports the "room for improvement" hypothesis (Halford & Bodenmann, 2013), which suggests that couples experiencing higher levels of distress may derive greater benefits from CRE than more satisfied couples (e.g., Bradford et al., 2017; Carlson et al., 2017; Halford et al., 2017; Quirk et al., 2014; Williamson et al., 2015, 2016). Yet the way researchers distinguished between distressed and nondistressed participants in these studies was often based on limited metrics or arbitrary cutoffs, such as using a single item or choosing seven out of 10 as a cutoff point.

In an effort to improve measurement of pretreatment relationship functioning, a recent study of moderators of CRE treatment effects employed a broader array of pretreatment relational characteristics, including measures of relationship happiness, commitment, and stability, and used a data-driven latent class analysis to select distressed versus nondistressed couples (Urganci et al., 2024). Results of this study revealed that benefits of CRE were predominantly seen in couples who exhibited higher relational functioning prior to the intervention. These findings raise questions about the best way to characterize pretreatment risk factors and indicate that relying on a single variable likely oversimplifies the experiences of the participants and fails to capture the intricate ways in which various aspects of relationship functioning interact with one another (Quintana, 2023).

While pretreatment relational dynamics are likely to be important in determining CRE treatment outcomes, the web of factors influencing treatment effects also extends to the broader context in which the couple lives. Basic studies of relationship functioning indicate that various risk factors, including personal characteristics and external stressors, are associated with relationship outcomes (Karney & Bradbury, 1995; McNulty et al., 2021), but these factors have been understudied in the relationship intervention literature. The few studies that have examined contextual risk factors suggest that these variables can moderate CRE treatment effects, but the findings remain mixed and incomplete. Recognizing that couples often face multiple, co-occurring challenges such as financial strain, mental health issues, and relational distress (Maisel & Karney, 2012), some researchers have adopted a cumulative risk approach when examining contextual stressors. This method involves summing the number of risk factors each couple presents with at baseline, with the assumption that the accumulation of risk exacerbates relationship difficulties and influences responsiveness to intervention (Rauer et al., 2008). Several studies using this approach have found that couples with more cumulative risk (e.g., those who were lower socioeconomic status, younger, or lacking external social support) tend to benefit more from CRE than lower risk couples, consistent with a "room for improvement" perspective in the contextual domain (Amato, 2014; Ritchie et al., 2023; Williamson et al., 2016).

However, even within this growing literature, the cumulative risk approach has limitations. By assigning equal weight to diverse risk factors and collapsing them into a single score, it assumes equifinality, which means that "individuals with very different risk factors or profiles may nevertheless experience similar relationship outcomes" (Rauer et al., 2008, p. 3). This assumption may obscure meaningful distinctions between qualitatively different profiles. For example, a minoritized couple experiencing high discrimination and low commitment and a White couple facing financial instability and emotional aggression might each receive a cumulative risk score of 2, yet the dynamics of their relationships and their potential response to CRE could be vastly different. Indeed, research suggests that not all risks are equal: In one study, couples with lower commitment improved more with CRE, while those facing aggression or problematic alcohol use actually fared worse than those in control conditions (Williamson et al., 2015). Thus, while cumulative risk models mark an important advance over univariate approaches, they still fall short of providing clear, actionable guidance for identifying which couples are most likely to benefit from CRE. A more flexible and nuanced analytic strategy is needed that can account for the distinctive and interactive effects of specific risk factors without forcing them into a single additive framework.

Overall, taking stock of the existing research on pretreatment characteristics, there is little consensus indicating which couples are more likely to benefit from CRE than others. Generally, a higher number of risk factors tends to be associated with receiving more benefit from the treatment, but the effects of specific risk factors are mixed, with some predicting improvement and others predicting poorer outcomes. Prior research often yields inconsistent or incomplete findings because it relies on simplified univariate or summative approaches that do not adequately capture the complex, distinct influences and interactions among multiple risk factors. Additionally, previous studies have relied on moderation in linear regression or subgroup analyses, methods that become difficult to interpret with high-dimensional data and require very large sample sizes which are often unattainable, resulting in underpowered studies unable to detect meaningful interaction effects (Quintana, 2023). Consequently, the field lacks clarity about which couples benefit most from CRE. Given the multifaceted nature of couples and the complexity of treatment response, more flexible, high-dimensional analytic approaches are needed to model multiple interacting factors simultaneously.

## The Present Study

The present study addresses these limitations by using causal forest analysis, which is a machine learning technique that was developed to assess treatment effects in RCTs in the presence of complex and high-dimensional covariates (Athey & Imbens, 2016; Wager & Athey, 2018). Causal forest is nonparametric, which means that it can estimate heterogeneous treatment effects without imposing a functional form on the predictors. For example, multiplicative interaction models that have been widely used to evaluate treatment effect heterogeneity assume the treatment effect changes at a constant rate with the moderator, an assumption that is not imposed in causal forest. This approach is well-suited to address this research question because it can accommodate a large array of parameters relative to sample size while mitigating the risk of overfitting by utilizing algorithms optimized for predictive power and accounting for the unique and interactive influence of each pretreatment risk factor.

In Study 1, we apply causal forest analysis to data from 6,298 couples who participated in an RCT of CRE to identify heterogeneity in treatment outcomes, then parse out the distinct and interactive roles that various pretreatment risks play in shaping the effectiveness of the CRE intervention. In Study 2, we test the generalizability and robustness of the results from Study 1 by applying the trained model to an independent data set of 1,595 couples who participated in an RCT of CRE. Study 1 is exploratory in nature, using the data-driven causal forest approach to explore how pretreatment characteristics contribute to heterogeneous treatment effects of CRE. Thus, we do not make hypotheses about the expected results of Study 1. In Study 2, we expect that the predictive algorithm developed in Study 1 will successfully predict treatment outcomes in these novel data.

# Study 1

## Method

### Transparency and Openness

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow Journal Article Reporting Standards (Appelbaum et al., 2018). The data reported in this article were obtained from publicly available data (the Supporting Healthy Marriage [SHM] evaluation, https://www.icpsr.umich.edu/web/DSDR/studies/34420). Analysis code and research materials are available at https://osf.io/42qyz. This study's design and its analysis were not preregistered.

### Participants and Procedure

Data are drawn from the SHM project, which is a RCT of CRE (Hsueh & Knox, 2014). Married[1] couples ($N$ = 6,298) who had or were expecting a child together and had a household income below $50,000[2] took part in the study, which was implemented at eight sites in seven different U.S. states. Enrollment occurred from February 2007 to December 2009. After providing informed consent, partners separately completed self-report questionnaires (T1) and then received their random assignment to the intervention condition or to the no-treatment control condition. A follow-up telephone interview was conducted separately with both partners about 12 months after enrollment (T2),[3] with a response rate of 80% for men and 85% for women. The analytic sample was comprised of individuals ($N$ = 9,977) who responded to the T1 and T2 surveys. This study used an intent-to-treat (ITT) approach, which includes all participants who were randomly assigned regardless of their uptake of the intervention.

### The SHM Program

The SHM program consisted of three parts: curriculum-based relationship and marriage education skills workshops in small groups, supplemental activities, and family support services. Local sites used one of four different curricula for their relationship skills workshops, all of which focused on common themes such as commitment, trust, conflict management, and promoting positive connections and intimacy. These four curricula offered 24–30 hr of programming, which local sites were free to deliver however they chose. For example, some sites chose to start participants with a full-day Saturday workshop, followed by weekly sessions, while others delivered the curriculum in a series of 9–15 weekly sessions. In addition to the relationship skills workshops, supplemental activities offered couples opportunities to attend educational events (e.g., seminars on financial management and parenting), participate in social events (e.g., date nights, family outings), practice skills from the workshops, and build networks with other couples in the program.

Finally, couples were paired with a family support staff member who had three goals: to maintain contact with couples to facilitate their participation in the other two program components, to help couples reduce family stressors and address family needs by linking them to community resources, and to reinforce key workshop themes in personal meetings with couples. Sessions were attended by both spouses; 83% of couples attended at least one workshop session, and couples received 60% of workshop hours, on average (17 hr). See Miller Gaubert et al. (2012) for additional details regarding recruitment, implementation, and intervention curricula.

[1] Although couples were required to be married at the time of enrollment, proof of marriage was not requested. Couples were asked to report their marital status at the 12-month assessment, where it was discovered that 80.9% of couples were married at the time of enrollment (Miller Gaubert et al., 2012).

[2] $60,000 for programs located in Seattle and the Bronx.

[3] SHM also included a 30-month follow-up. However, we do not use these data because the data set used in Study 2 only has a 12-month follow-up.

### Measures

**Dependent Variables.** Four treatment outcome variables were tested, consistent with the primary outcomes used in the original SHM evaluation (Lundquist et al., 2014).

***Relationship Happiness.*** The participant's global appraisal of the relationship was assessed with a single item asking, "All things considered, on a scale from 1 to 7, where 1 is *completely unhappy* and 7 is *completely happy*, how happy are you with your marriage to SPOUSENAME?"

***Positive Communication Skills.*** Seven items based on the Gottman Sound Relationship House Questionnaire were used to measure how well the couple communicates during disagreements. An example item is "We are good at working out our differences." Items were scored on a 1 to 4 scale, where 1 = *never* and 4 = *often*. Responses were averaged to form the scale score ($\alpha$ = .79).

***Negative Emotions and Behaviors.*** Seven items were used to measure negative couple interactions during disagreements. An example item is "My spouse was rude and mean to me when we disagreed." Items were scored on a 1 to 4 scale, where 1 = *never* and 4 = *often*. Responses were averaged to form the scale score ($\alpha$ = .90).

***Relationship Stability.*** Whether the relationship was still intact at the 12-month follow-up was assessed with a single item. Respondents were asked, "What is your marriage status?" with response options of married (living together), married (living apart), divorced, and separated. If *married* was selected for both partners, a score of 1 was given. Otherwise, a score of 0 was given.

**Pretreatment Characteristics.** All available baseline variables (31 in total) were used as pretreatment characteristics. These included race, income, religiosity, education, psychological distress, substance abuse, social support, adverse childhood experiences, stressful life events, perceived stress, relationship happiness, thoughts that the relationship was in trouble, and relational aggression measured at the individual level (i.e., self-reported and partner-reported are both included in the model); relationship length, cohabitation history, whether this was a remarriage, and number of children measured at the couple level; and one dummy variable to indicate gender because all couples are male/female dyads. A detailed description of how each of these variables was measured is available in the Supplemental Material.

### Analytic Approach

All analyses were performed in R (Version 4.2.2; R Core Team, 2024) using the "grf" package for causal forest analyses (Tibshirani et al., 2022). The causal forest algorithm extends the random forest framework to the causal setting by developing decision trees that partition the data based on covariates that explain differences in treatment effect, rather than differences in outcomes alone (as is the case in random forest; see Athey & Imbens, 2016). Each tree identifies subgroups with similar covariate profiles and estimates localized treatment effects within those subgroups.

Following guidance from Wager and Athey (2018), we instructed the algorithm to construct 5,000 causal trees using bootstrapped random subsamples and 25 randomly selected pretreatment characteristics. To account for the clustered nature of the data, we incorporated individual-level pretreatment characteristics from both partners of each couple simultaneously into the models. Additionally, we implemented a sampling strategy within the algorithm to ensure balanced representation across sites and to prevent double counting of couples within each tree. Specifically, the algorithm was configured to draw equally from each site and to select only one partner from each couple for each tree. This approach maintains the integrity of the couple-level and site-level data structure while allowing for a comprehensive analysis of individual-level effects within the context of the relationship. The algorithm then assembled all the trees to estimate the conditional average treatment effect (CATE), which identifies how much the treatment is expected to affect a specific individual, given their characteristics. It is these individualized estimates produced by causal forest that allow for fine-grained analysis of effect modifiers compared to the traditional analytic approach to RCTs, which compares the mean levels of the treatment and control groups and produces only sample-level estimates.

A two-stage causal forest approach was used to obtain less biased estimates, utilizing the honest splitting technique (Athey & Imbens, 2016). Through recursive partitioning, the data were first divided into two subsets for each tree. The first subset is the training data set, which was used to construct causal trees, with the goal to maximize heterogeneity in treatment effects across leaves. The second subset is the estimation data set, which was used to estimate the treatment effects of each leaf of the trees. When estimating individual CATEs, the algorithm employed an augmented inverse-propensity weighted estimator in each leaf. Augmented inverse-propensity weighting incorporates the treatment assignment and the propensity of being treated given pretreatment characteristics to ensure precise estimations of CATEs. This method demonstrates robustness in estimating treatment effects even in the presence of attrition (Kurz, 2022).

After estimating CATEs, the next step is to determine whether the treatment effects are heterogeneous. Heterogeneous treatment effects occur when the effect of a treatment or intervention is not constant across all individuals but instead varies systematically with individual characteristics. In contrast, if treatment effects are not heterogeneous, this indicates that the treatment effect does not systematically vary based on any observed characteristics or subgroups. To test for the presence of heterogeneous treatment effects, the methodology defined by Chernozhukov et al. (2018) was employed using a best linear prediction test characterized by $\gamma$ (average treatment effect [ATE]) and $\beta$ (differential treatment effect), where a significant $\beta$ indicates that there is significant heterogeneity around the ATE.

For the models that evidenced significant heterogeneity in treatment effects, the covariates (i.e., pretreatment characteristics) were analyzed to determine which ones contributed significantly to that heterogeneity. This is represented through a variable importance value, which indicates the extent to which each pretreatment characteristic influences the treatment effects by calculating what percentage of nodes in the causal trees are generated by a given predictor. Importance values are weighted by the order of the split, where higher order splits indicate higher weights.

After estimating the importance of each pretreatment characteristic, we used feature selection models to identify the constellation of pretreatment characteristics that were most predictive of treatment outcomes (Cai et al., 2018). In these models, the top 10 most important pretreatment characteristics were entered using a backward stepwise approach by iteratively fitting causal forest models with subsets of pretreatment characteristics. At each iteration, we removed the least important variable and reevaluated whether statistically significant heterogeneity persisted. We continued the fitting process until we identified the minimal set of pretreatment characteristics that retained significant heterogeneity.

**Table 1**
*Descriptive Statistics of All Variables*

| Variable | Study 1 | | Study 2 | |
|---|---|---|---|---|
| | Mean/prop | *SD* | Mean/proportion | *SD* |
| Female | 50.0% | | 50.0% | |
| Race | | | | |
| White | 20.5% | | 23.2% | |
| Black | 11.3% | | 24.5% | |
| Hispanic | 43.43% | | 44.4% | |
| Multi/other | 24.8% | | 7.9% | |
| Annual income (U.S. dollar) | 14,434 | 12,774 | 20,612 | 21,999 |
| Religiosity | 2.55 | 1.11 | | |
| Education (years) | 11.83 | 2.54 | 11.76 | 2.83 |
| Psychological distress | 1.13 | 0.85 | 1.73 | 0.66 |
| Substance abuse | 0.40 | 0.99 | | |
| Social support | 0.82 | 0.87 | | |
| Adverse childhood experiences | 1.89 | 0.90 | | |
| Stressful events | 1.82 | 1.37 | 0.81 | 0.95 |
| Perceived stress | 1.95 | 0.74 | | |
| Relationship in trouble | 56% | | 66% | |
| Relational aggression | 2.32 | 0.91 | 0.58 | 1.07 |
| Relationship length (years) | 5.14 | 3.74 | 9.70 | 7.18 |
| Cohabiting history | 67% | | | |
| Remarried | 19% | | 20% | |
| Number of children | 2.09 | 1.14 | 2.07 | 1.18 |
| Baseline relationship happiness | 5.42 | 1.60 | 7.32 | 2.17 |
| Relationship happiness T2 | 5.88 | 1.25 | 7.77 | 2.34 |
| Relationship stability T2 | 81% | | | |
| Positive communication skills T2 | 3.20 | 0.57 | | |
| Negative emotions and behaviors T2 | 2.17 | 0.78 | 2.25 | 0.92 |

*Note*. $N$ = 6,298 couples (12,596 individuals) for Study 1 and $N$ = 1,595 couples (3,190 individuals) for Study 2. Not all variables are measured on the same scale across studies, so values should not be directly compared.

This process allowed us to identify the most parsimonious set of pretreatment characteristics needed to accurately predict how a couple will respond to CRE.

Finally, although importance values indicate which pretreatment characteristics contribute to heterogeneity in treatment outcomes, this metric does not characterize the direction of the association between the pretreatment characteristic and treatment outcome. To understand the pretreatment characteristics of the individuals who benefitted most and least from the intervention, we conducted a quantile-based subgroup analysis in which individuals were ranked by their estimated treatment effect and grouped into quartiles. We computed the ATE and the distribution of pretreatment characteristics within each quartile and summarized covariate patterns using descriptive statistics to characterize the participants within each group.[4]

## Results

### *Descriptive Statistics*

Participants were racially and ethnically diverse, with 43% of couples identifying as both Hispanic, 21% identifying as both White, 11% identifying as both Black, and 25% of couples identifying as another race/ethnicity or couples who differ in racial/ethnic background. Couples had two children on average and had been married an average of 5 years. There was a range of relationship functioning, with 56% of individuals reporting that they thought their relationship was in trouble. The average level of relationship happiness at baseline was 5.4 ($SD = 1.60$) on a 1 to 7 scale. Table 1 presents descriptive statistics for all baseline characteristics.

### *Estimation of ATEs*

A separate causal forest model was estimated for each of the four treatment outcome variables. We first compared the ATEs estimated by the causal forest models to the treatment effects from the original evaluation of the SHM program (Hsueh et al., 2012) to confirm that the causal forest models had successfully estimated the treatment effects. In the original evaluation, the intervention had an impact of .150 on relationship happiness ($p < .10$, $d = .13$),[5] and the causal forest model estimated the ATE as .159. The original evaluation found a treatment effect of .05 for women and .07 for men for positive communication skills ($p < .10$, $d = .08$ and $p < .10$, $d = .11$, respectively), and the causal forest model estimated the ATE for positive communication skills as .059. The original evaluation found a treatment effect of −.07 for men and −.09 for women for negative emotions and behaviors ($p < .10$, $d = -.08$ and $p < .10$, $d = -.12$, respectively), and the causal forest model estimated the ATE for positive communication skills as −.086. Finally, in the original

[4] Further analyses were conducted to assess the independent predictive validity of the top important pretreatment characteristics using the Rank Average Treatment Effect framework and partial dependence plots. The results of these analyses are presented in the Supplemental Material.

[5] The SHM report does not provide exact $p$ values for these analyses.

evaluation, the treatment effect on relationship stability was .008 ($p > .10$, $d = .00$), and the causal forest model estimated the ATE for relationship stability as .007. Overall, the ATEs derived from the causal forest models were very similar to the treatment effects reported in the original evaluation.

### *Estimation of Heterogeneity in Treatment Effects*

Figure 1 shows the distribution of CATEs estimated by causal forest for each outcome, where the red vertical line represents the ATEs. Examining these figures shows that some appear to be not normally distributed, which suggests heterogeneity in treatment effects. We then used best linear predictions to confirm if significant heterogeneity was present in any of the outcomes. Treatment effects on relationship happiness ($\beta = 1.452$, $p = .014$) and negative emotions and behaviors ($\beta = 1.600$, $p = .007$) had significant heterogeneity, indicating that the effect of the treatment on these two outcomes differed systematically based on individual characteristics. Treatment effects on relationship stability ($\beta = -4.425$, $p = .995$) and positive communication skills ($\beta = 1.103$, $p = .107$) did not have significant heterogeneity, indicating that the treatment effect on these two outcomes did not systematically vary based on any observed characteristics. Thus, for the remainder of the results, we focus on identifying the pretreatment characteristics that contributed to treatment effects on relationship happiness and negative emotions and behaviors.

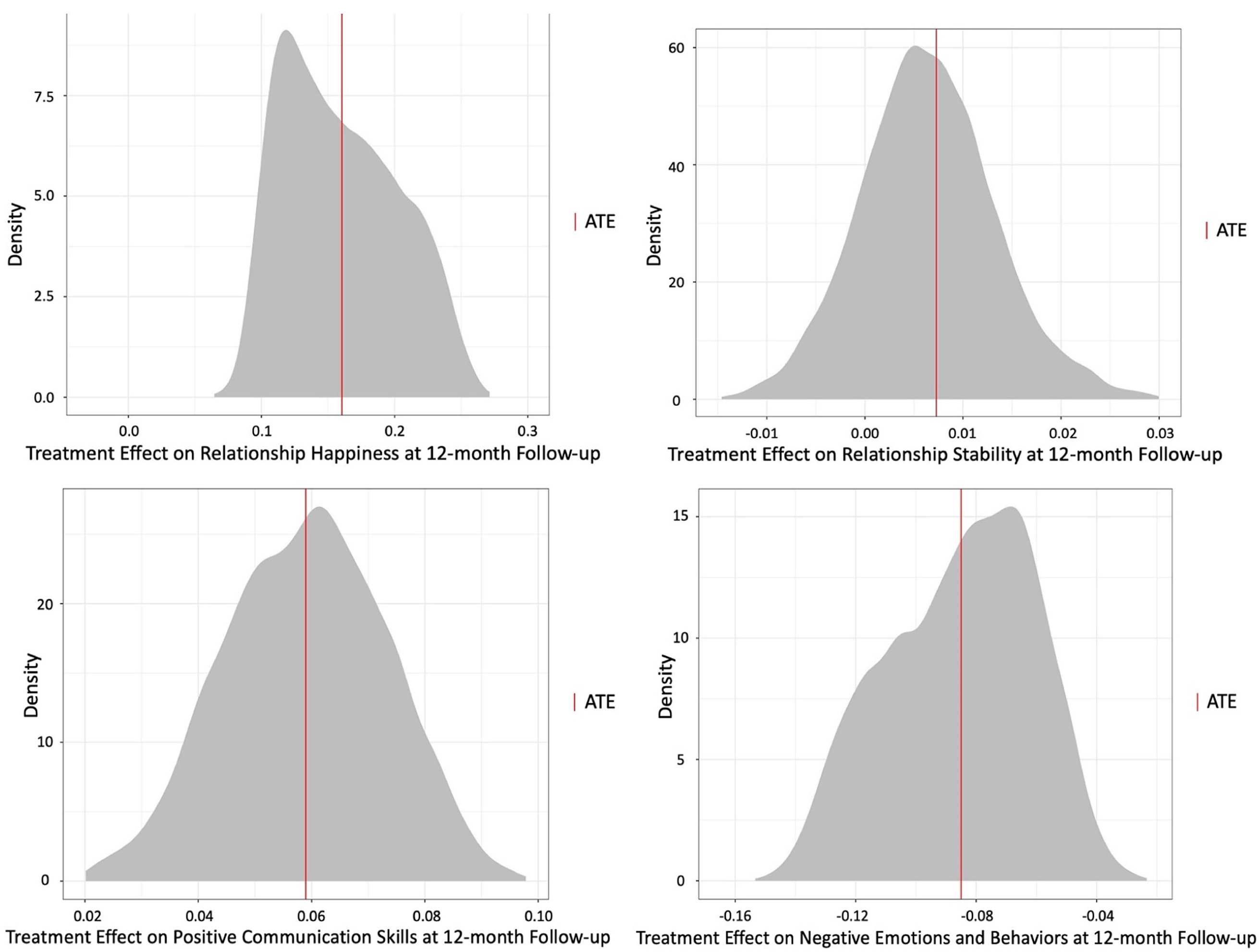
### *Moderators of Relationship Happiness*

We used variable importance derived from the causal forest models to determine which pretreatment characteristics contributed to the heterogeneous treatment effect on relationship happiness (see Table 2). Baseline psychological distress was the most important variable, contributing to 16.4% of the splits in the causal forest. This was followed by baseline relationship happiness, which contributed to 7.7% of the splits, and partner-reported psychological distress, which contributed to 6.2% of the splits.

We next conducted a feature selection model, which identified a set of six pretreatment characteristics that made up the most parsimonious model for predicting treatment effects on relationship happiness

**Figure 1**

*Distributions of Conditional Average Treatment Effects for Each of the Outcome Variables in Study 1*




*Note*. ATE = average treatment effect. See the online article for the color version of this figure.

**Table 2**
*Variable Importance Values for Causal Forest Models Predicting Relationship Happiness in Study 1*

| | Importance value | |
|---|---|---|
| Pretreatment characteristic | Full model | Parsimonious model |
| Psychological distress | 0.164 | 0.269 |
| Baseline relationship happiness | 0.077 | 0.178 |
| Partner psychological distress | 0.064 | 0.158 |
| Partner adverse childhood experiences | 0.057 | 0.145 |
| Adverse childhood experiences | 0.051 | 0.135 |
| Perceived stress | 0.048 | 0.116 |
| Relational aggression | 0.046 | |
| Partner relational aggression | 0.042 | |
| Partner baseline relationship happiness | 0.036 | |
| Relationship in trouble | 0.034 | |

*Note.* $N$ = 9,351. For the full model, only the top 10 most important variables are presented.

(see Table 2). Psychological distress once again emerged as the most influential predictor of treatment effect heterogeneity (26.9%), followed by baseline relationship happiness (17.8%), partner-reported psychological distress (15.8%), partner-reported adverse childhood experiences (14.5%), adverse childhood experiences (13.5%), and perceived stress (11.6%).

Finally, we grouped participants into quartiles based on their CATE and summarized the pretreatment characteristics of these groups. As shown in Table 3, participants who benefitted the most from the intervention (i.e., those in the top quartile) had an ATE of $d$ = .28. These participants were generally characterized by low levels of relationship functioning (e.g., low relationship happiness, high relational aggression, 94% thought their relationship was in trouble), high levels of individual distress (e.g., high psychological distress, high perceived stress), and high contextual stress (e.g., high adverse childhood experiences, high stressful life events). On the opposite end, participants who benefitted least (i.e., those in the bottom quartile) had an ATE of $d$ = .05. These participants were generally characterized by high levels of relationship functioning (e.g., only 7% thought their relationship was in trouble), low individual distress, and low contextual stress.

### *Moderators of Negative Emotions and Behaviors*

The variable importance values for negative emotions and behaviors (presented in Table 4) indicate that baseline psychological distress was the most influential factor, contributing to 16.5% of the splits in the causal forest. This was followed by perceived stress, which contributed to 6.0% of the splits, and partner-reported psychological distress, which contributed to 5.5% of the splits.

We next conducted a feature selection model, which identified three pretreatment characteristics as part of the most parsimonious model for predicting treatment effects on negative emotions and behaviors (Table 4). Psychological distress was again the most influential characteristic (41.5%), followed by partner-reported psychological distress (29.3%) and educational attainment (29.2%).

Finally, we grouped participants into quartiles based on their CATE and summarized the pretreatment characteristics of these groups. As shown in Table 5, participants who benefited the most from the intervention (i.e., those in the top quartile) had an ATE of $d$ = .20. Similar to the results for relationship happiness, these participants were generally characterized by low levels of relationship functioning (e.g., low relationship happiness, high relational aggression, 88% thought their relationship was in trouble), high levels of individual distress (e.g., high psychological distress, high perceived stress), and high contextual stress (e.g., high adverse childhood experiences, high stressful life events). On the opposite end, participants who benefitted least (i.e., those in the bottom quartile) had an ATE of $d$ = .04. These participants were generally characterized by high levels of relationship functioning (e.g., only 21% thought their relationship was in trouble), low individual distress, and low contextual stress.

# Study 2

Study 2 was designed to assess the robustness of the results from Study 1. Study 1 indicated which pretreatment characteristics carry the most weight in determining treatment effects, but if those results are idiosyncratic to the data in Study 1, then the model will not be able to accurately estimate treatment effects in a new data set. However, if the characteristics that were most predictive of treatment effects in Study 1 are also important in Study 2, the trained models will accurately predict ATEs, and the pattern of baseline characteristics will be similar to that observed in Study 1. In other words, here we sought to test whether new information about pretreatment characteristics can be plugged into the algorithms developed in Study 1 to accurately predict treatment outcomes in a novel CRE intervention.

## Method

### *Transparency and Openness*

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study, and we follow Journal Article Reporting Standards (Appelbaum et al., 2018). Data and materials are available through the Inter-university Consortium for Political and Social Research data repository (Study No. 37843). Analysis code is available at https://osf.io/42qyz. This study's design and its analysis were not preregistered.

### *Participants and Procedure*

Data are drawn from the Parents and Children Together (PACT) project, which is a RCT of CRE (Moore et al., 2018). Different-gender couples ($N$ = 1,595) aged 18 or above who were expecting or parenting a child together took part in the study, which was implemented across two sites (El Paso, Texas, and Bronx, New York). Enrollment occurred from July 2013 through April 2015. After providing informed consent, partners separately completed a 30-min baseline phone survey (T1) and then received their random assignment to the intervention condition or the no-treatment control condition. Telephone interviews were conducted separately with both partners about 12 months after enrollment (T2), with a response rate of 91% for women and 85% for men. The analytic sample was comprised of individuals who responded to the T2 outcomes

**Table 3**
*Means and Standard Deviations of Quantile Grouping for Relationship Happiness in Study 1*

| Pretreatment characteristic | Top 25% | 26%–50% | 51%–75% | Bottom 25% |
|---|---|---|---|---|
| Psychological distress | 2.07 (.70) | 1.18 (.61) | 0.68 (.44) | 0.38 (.33) |
| Baseline relationship happiness | 4.16 (1.60) | 5.27 (1.47) | 6.00 (1.16) | 6.65 (.55) |
| Partner psychological distress | 1.47 (.88) | 1.18 (.83) | 0.92 (.75) | 0.77 (.69) |
| Partner adverse childhood experiences | 1.93 (.87) | 1.89 (.89) | 1.80 (.87) | 1.81 (.91) |
| Adverse childhood experiences | 2.17 (.96) | 1.92 (.90) | 1.73 (.83) | 1.62 (.76) |
| Perceived stress | 2.58 (.66) | 2.07 (.56) | 1.69 (.55) | 1.30 (.41) |
| Relational aggression | 2.78 (.84) | 2.41 (.85) | 2.11 (.82) | 1.74 (.73) |
| Partner relational aggression | 2.81 (.84) | 2.42 (.88) | 2.11 (.82) | 1.78 (.72) |
| Partner baseline relationship happiness | 4.61 (1.66) | 5.29 (1.55) | 5.85 (1.37) | 6.33 (1.05) |
| Relationship in trouble | 94.1% | 70.5% | 38.3% | 7.3% |
| Partner perceived stress | 2.28 (.72) | 2.03 (.72) | 1.80 (.69) | 1.55 (.57) |
| Income | 12,688 (12,385) | 14,555 (12,821) | 15,171 (12,924) | 16,351 (13,238) |
| Partner income | 15,087 (12,562) | 14,670 (12,795) | 15,217 (13,093) | 14,655 (13,198) |
| Partner stressful events | 2.04 (1.43) | 1.84 (1.35) | 1.64 (1.31) | 1.49 (1.20) |
| Education | 11.94 (2.52) | 11.90 (2.56) | 12.03 (2.63) | 11.72 (2.63) |
| Stressful events | 2.12 (1.39) | 1.90 (1.37) | 1.62 (1.29) | 1.36 (1.14) |
| Partner education | 11.86 (2.48) | 11.84 (2.59) | 11.92 (2.65) | 11.87 (2.64) |
| Partner race | | | | |
| White | 19.5% | 18.9% | 19.9% | 25.6% |
| Black | 7.5% | 10.7% | 11.0% | 14.8% |
| Hispanic | 48.4% | 43.0% | 44.9% | 39.5% |
| Social support | 1.04 (.83) | 0.86 (.86) | 0.72 (.88) | 0.65 (.87) |
| Partner substance abuse | 0.58 (1.23) | 0.45 (1.0) | 0.28 (.76) | 0.16 (.57) |
| Relationship length | 5.71 (3.71) | 5.31 (3.81) | 5.19 (3.78) | 4.97 (3.75) |
| Substance abuse | 0.58 (1.19) | 0.41 (1.0) | 0.26 (.77) | 0.17 (.56) |
| Number of children | 2.29 (1.14) | 2.19 (1.15) | 2.05 (1.14) | 1.96 (1.12) |
| Partner religiosity | 2.46 (.023) | 2.52 (.023) | 2.60 (.023) | 2.87 (.023) |
| Remarried | 23.4% | 19.6% | 16.7% | 13.6% |
| Religiosity | 2.48 (1.08) | 2.52 (1.10) | 2.62 (1.12) | 2.86 (1.12) |
| Race | | | | |
| White | 19.5% | 18.9% | 19.9% | 25.6% |
| Black | 7.5% | 10.7% | 11.0% | 14.8% |
| Hispanic | 48.8% | 43.0% | 44.9% | 39.5% |
| Partner social support | 0.94 (.85) | 0.81 (.86) | 0.75 (.88) | 0.75 (.90) |
| Partner relationship in trouble | 81.5% | 63.3% | 44.1% | 21.8% |
| Cohabitation | 71.5% | 68.2% | 63.2% | 55.8% |
| Female | 57.3% | 52.0% | 49.6% | 44.9% |
| ATE[a] | 0.32 (.02) | 0.18 (.01) | 0.08 (.01) | 0.06 (.01) |
| Effect size[b] (*d*) | 0.28 | 0.15 | 0.07 | 0.05 |

*Note.* $N$ = 9,351. Standard errors are presented in parentheses after the mean. ATE = average treatment effect.
[a] ATE calculated from causal forest. [b] Calculated using *SD* from the original report.



($N$ = 2,801 for relationship happiness and $N$ = 2,747 for negative emotions and behaviors). This study used an ITT approach, which includes all participants who were randomly assigned regardless of their uptake of the intervention.

## The PACT Program

The PACT program consisted of three primary components: curriculum-based relationship and marriage education skills workshops in small groups, ancillary job and career advancement services, and family support services. Sites used one of two curricula for their relationship skills workshops, both of which focused on common themes such as strategies to avoid conflict, providing support, and effective communication. The curricula offered 18–24 hr of programming, which local sites were free to deliver however they chose.

Overall, 87% of couples attended at least one session of relationship education, and 68% of couples attended at least half the sessions. All intervention programming (as well as the baseline and follow-up surveys) was available in English and Spanish.

In addition to the core relationship education programming, additional job and career advancement services were offered to participants, including 2-hr stand-alone employment workshops covering topics such as preparing resumes and developing soft skills, and one-on-one services from an employment specialist. Furthermore, one site also integrated economic and financial well-being topics into their relationship education workshops. Overall, these services were of fairly low intensity and did not have a strong uptake (see Zaveri & Baumgartner, 2016). Finally, treatment group participants were also offered individual case management services and relationship education booster sessions throughout the 1-year period.

## Measures

**Dependent Variables.** Two treatment outcome variables were identified that matched the two treatment outcomes with significant heterogeneity in Study 1.

**Table 4**
*Variable Importance Values for Causal Forest Models Predicting Negative Emotions and Behaviors in Study 1*

| Pretreatment characteristic | Importance value | |
|---|---|---|
| | Full model | Parsimonious model |
| Psychological distress | 0.165 | 0.415 |
| Perceived stressed | 0.060 | |
| Partner psychological distress | 0.055 | 0.293 |
| Baseline relationship happiness | 0.053 | |
| Relational aggression | 0.050 | |
| Educational attainment | 0.049 | 0.292 |
| Partner adverse childhood experiences | 0.049 | |
| Number of children | 0.045 | |
| Adverse childhood experiences | 0.043 | |
| Partner stressful life events | 0.039 | |

*Note*. $N$ = 9,977. For the full model, only the top 10 most important variables are presented.

***Relationship Happiness.*** The participant's global appraisal of the relationship was assessed with a single item asking, "On a scale from 0 to 10, where 0 is *not at all happy* and 10 is *completely happy*, taking all things together, how happy are you with PARTNER?"

***Negative Emotions and Behaviors.*** Seven items were used to measure negative couple interactions during disagreements. An example item is "My spouse was rude and mean to me when we disagreed." Items were scored on a 1 to 4 scale, where 1 = *never* and 4 = *often*. Responses were averaged to form the scale score (α = .93).

**Pretreatment Characteristics.** We identified variables in the PACT data that matched the pretreatment characteristics used in Study 1. Out of 31 pretreatment characteristic variables that were used in Study 1, we identified matches for 20 in the PACT data. Variables that were available for inclusion in analyses were gender, race, income, education, psychological distress, stressful life events, relationship happiness, thoughts that the relationship was in trouble, relational aggression, relationship length, whether this was a remarriage, and number of children. Religiosity, substance abuse, social support, adverse childhood experiences, perceived stress, and cohabitation history were not available in the PACT data. A detailed description of how each of these variables was measured is available in the Supplemental Materials.

### *Analytic Approach*

Analyses were performed in R using the "grf" package (Version 4.2.2; R Core Team, 2024; Tibshirani et al., 2022). Given the unobservable nature of individual treatment effects and the lack of formal testable metrics for validating these models on new data sets (Gong et al., 2021; Shiba & Inoue, 2024), we employed a series of techniques to assess the performance and generalizability of the causal forest models derived in Study 1. We first compared the ATEs estimated by the trained causal forest models to the treatment effects from the original evaluation of the PACT program to confirm that the causal forest models had successfully estimated the treatment effects (Moore et al., 2018). Next, consistent with Study 1, we conducted a quantile-based subgroup analysis in which individuals were ranked by their estimated treatment effect and grouped into quartiles. We computed the ATE and the distribution of pretreatment characteristics within each quartile and summarized covariate patterns using descriptive statistics to determine whether the characteristics of the participants within each quartile group were similar to those in Study 1. This approach, while not definitive, allowed us to gauge whether the patterns of association between pretreatment characteristics and treatment effects were consistent with what we observed in Study 1.

## Results

Table 1 presents the baseline characteristics of participants in Study 2. The participants in PACT were generally comparable to the participants in SHM on demographics and relationship functioning, though there was a larger proportion of Black couples in PACT.

### *Relationship Happiness*

We first examined whether the fully trained model from Study 1 could accurately estimate ATEs on relationship happiness in these new data. The full model estimated an ATE of 0.143 in the PACT data, which is very similar to the treatment effect reported in the original evaluation (impact = 0.15, $p$ = .12, $d$ = .07; Moore et al., 2018). We next applied the most parsimonious model identified through the feature selection process in Study 1 to estimate the ATE. The parsimonious model estimated the treatment effect as 0.155, which again is very close to that from the original evaluation. Supplemental Figures S5 and S6 present the distribution of estimated conditional treatment effects for these two models.

Next, we grouped participants into quartiles based on their CATE and summarized the pretreatment characteristics of these groups. As shown in Table 6, participants who benefitted the most from the intervention (i.e., those in the top quartile) had an ATE of $d$ = .08. These participants were generally characterized by low levels of relationship functioning (e.g., low relationship happiness, high relational aggression, 99% thought their relationship was in trouble), high levels of individual distress (e.g., high psychological distress), and high contextual stress (e.g., high stressful life events). On the opposite end, participants who benefitted least (i.e., those in the bottom quartile) had an ATE of $d$ = .05. These participants were generally characterized by high levels of relationship functioning (e.g., only 12% thought their relationship was in trouble), low individual distress, and low contextual stress. This pattern of results is consistent with those obtained in Study 1.

### *Negative Emotions and Behaviors*

The trained causal forest model for negative emotions and behaviors estimated an ATE of −0.074 in the PACT data, which is similar to the treatment effect reported in the original evaluation (impact = 0.05, $p$ = .09, $d$ = .07). We next applied the most parsimonious model identified through the feature selection process in Study 1 to estimate the ATE. The parsimonious model estimated the treatment effect as 0.106, which was twice as large as that from the original report. This result suggests that this three-variable parsimonious model was less successful at accurately estimating treatment effects. Supplemental Figures S5 and S6 present the distribution of conditional treatment effects for these two models.

Next, we grouped participants into quartiles based on their CATE and summarized the pretreatment characteristics of these groups. As shown in Table 7, participants who benefitted the most from the intervention (i.e., those in the top quartile) had an ATE of $d$ = .13. These participants were generally characterized by low levels of

**Table 5**
*Mean and Standard Errors of Quantile Groupings for Negative Emotions and Behaviors in Study 1*

| Pretreatment characteristic | Top 25% | 26%–50% | 51%–75% | Bottom 25% |
|---|---|---|---|---|
| Psychological distress | 2.12 (.69) | 1.18 (.58) | 0.66 (.44) | 0.43 (.34) |
| Perceived stress | 2.61 (.66) | 2.06 (.58) | 1.70 (.57) | 1.36 (.45) |
| Partner psychological distress | 1.54 (.89) | 1.18 (.84) | 0.96 (.75) | 0.74 (.65) |
| Baseline relationship happiness | 4.24 (1.66) | 5.26 (1.48) | 5.93 (1.23) | 6.46 (.86) |
| Partner relational aggression | 2.83 (.83) | 2.47 (.89) | 2.17 (.83) | 1.75 (.71) |
| Education | 12.10 (2.56) | 12.11 (2.66) | 11.98 (2.60) | 11.29 (2.24) |
| Partner adverse childhood experiences | 1.87 (.87) | 1.94 (.91) | 1.86 (.89) | 1.83 (.91) |
| Number of children | 2.42 (1.15) | 2.26 (1.17) | 2.17 (1.13) | 1.66 (1.0) |
| Adverse childhood experiences | 2.17 (.95) | 1.98 (.91) | 1.75 (.82) | 1.59 (.77) |
| Partner stressful events | 2.27 (1.42) | 2.02 (1.37) | 1.82 (1.36) | 1.62 (1.20) |
| Partner income | 15,861 (12,588) | 15,297 (13,074) | 15,399 (12,971) | 12,571 (12,613) |
| Religiosity | 2.64 (1.06) | 2.65 (1.11) | 2.69 (1.12) | 2.42 (1.13) |
| Relational aggression | 2.71 (.85) | 2.41 (.87) | 2.19 (.83) | 1.89 (.82) |
| Income | 12,947 (12,522) | 14,492 (12,905) | 14,534 (12,961) | 16,551 (12,925) |
| Stressful events | 2.04 (1.38) | 1.82 (1.37) | 1.63 (1.32) | 1.65 (1.26) |
| Partner baseline relationship happiness | 4.68 (1.68) | 5.29 (1.56) | 5.73 (1.46) | 6.18 (1.18) |
| Partner education | 11.89 (2.50) | 11.93 (2.61) | 12.00 (2.62) | 11.61 (2.48) |
| Partner substance abuse | 0.58 (1.16) | 0.45 (1.05) | 0.31 (.86) | 0.20 (.64) |
| Relationship length | 5.94 (3.75) | 5.63 (3.79) | 5.34 (3.78) | 4.05 (3.46) |
| Substance abuse | 0.48 (1.09) | 0.42 (1.02) | 0.28 (.81) | 0.27 (.74) |
| Partner perceived stress | 2.27 (.72) | 2.02 (.73) | 1.82 (.69) | 1.62 (.61) |
| Relationship in trouble | 88.1% | 67.3% | 41.4% | 20.5% |
| Social support | 1.01 (.83) | 0.86 (.87) | 0.76 (.89) | 0.69 (.88) |
| Partner religiosity | 2.60 (1.07) | 2.59 (1.13) | 2.68 (1.13) | 2.49 (1.13) |
| Partner social support | 0.90 (.85) | 0.78 (.85) | 0.74 (.88) | 0.86 (.91) |
| Remarried | 23.3% | 19.5% | 17.3% | 14.7% |
| Partner race | | | | |
| White | 17.9% | 19.1% | 20.4% | 27.1% |
| Black | 50.6% | 45.1% | 42.5% | 35.2% |
| Hispanic | 9.5% | 11.5% | 12.6% | 11.4% |
| Race | | | | |
| White | 17.9% | 19.1% | 20.4% | 27.1% |
| Black | 50.6% | 45.1% | 42.5% | 35.2% |
| Hispanic | 9.5% | 11.5% | 12.6% | 11.4% |
| Female | 61.2% | 54.1% | 52.2% | 37.8% |
| Cohabitation | 70.2% | 65.7% | 62.7% | 63.0% |
| Partner relationship in trouble | 79.5% | 61.9% | 46.3% | 29.6% |
| ATE[a] | −0.16 (.01) | −0.09 (.01) | −0.07 (.01) | −0.03 (.01) |
| Effect size[b] (*d*) | 0.20 | 0.11 | 0.09 | 0.04 |

*Note.* $N = 9{,}977$. Standard errors are presented in parentheses after the mean. ATE = average treatment effect.
[a] ATE calculated from causal forest. [b] Calculated using *SD* from the original report.



relationship functioning (e.g., low relationship happiness, high relational aggression, 95% thought their relationship was in trouble), high levels of individual distress (e.g., high psychological distress), and high contextual stress (e.g., high stressful life events). On the opposite end, participants who benefitted least (i.e., those in the bottom quartile) had an ATE of $d = .08$. These participants were generally characterized by high levels of relationship functioning (e.g., only 12% thought their relationship was in trouble), low individual distress, and low contextual stress. This pattern of results is consistent with those obtained in Study 1.

## Discussion

Using a data-driven machine learning method, the present study advances our understanding of which couples benefit most and least from CRE. Results indicated that there was heterogeneity in the extent to which the treatment impacted 12-month outcomes on relationship happiness and negative emotions and behaviors, and this heterogeneity could be reliably predicted by couples' pretreatment characteristics.

Specifically, results indicated that participants who were experiencing higher levels of distress, both individually and relationally, benefited the most from CRE. This general pattern of results, in which those with lower levels of pretreatment functioning reap the most from CRE, aligns with a substantial body of previous research. However, the specific constructs that were most predictive of treatment outcomes diverged somewhat from those that have been examined in past literature. The vast majority of past research has focused on characteristics of the relationship as moderators of treatment outcomes (e.g., level of relationship satisfaction when entering the program). Indeed, a large literature has demonstrated that pretreatment relationship distress is positively associated with intervention outcomes, with more distressed couples showing greater improvements (e.g., Bradford et al., 2017; Carlson et al., 2017; Halford et al., 2017; Quirk et al., 2014; Williamson et al., 2015, 2016). Results of the present study were consistent with this body of work, indicating that baseline relationship happiness was one of

**Table 6**
*Means and Standard Deviations of Quantile Grouping for Relationship Happiness in Study 2*

| Pretreatment characteristic | Top 25% | 26%–50% | 51%–75% | Bottom 25% |
|---|---|---|---|---|
| Psychological distress | 2.46 (.66) | 1.78 (.48) | 1.51 (.37) | 1.17 (.22) |
| Baseline relationship happiness | 5.34 (2.35) | 7.10 (1.70) | 8.01 (1.46) | 8.95 (1.07) |
| Partner psychological distress | 1.97 (.74) | 1.73 (.65) | 1.65 (.62) | 1.57 (.56) |
| Relational aggression | 1.18 (1.38) | 0.62 (1.04) | 0.34 (.78) | 0.11 (.43) |
| Partner relational aggression | 1.15 (1.38) | 0.62 (1.08) | 0.36 (.81) | 0.19 (.57) |
| Partner baseline relationship happiness | 5.92 (2.40) | 7.22 (1.94) | 7.89 (1.81) | 8.45 (1.48) |
| Relationship in trouble | 98.8% | 92.4% | 58.9% | 12.1% |
| Income | 19,416 (20,568) | 20,448 (27,732) | 21,621 (23,448) | 21,073 (17,556) |
| Partner income | 23,352 (26,760) | 21,456 (27,324) | 19,440 (16,188) | 18,264 (17,076) |
| Partner stressful events | 0.99 (1.06) | 0.85 (0.95) | 0.70 (0.90) | 0.69 (0.86) |
| Education | 11.91 (2.77) | 11.95 (2.79) | 11.89 (2.83) | 11.28 (2.88) |
| Stressful events | 1.15 (1.05) | 0.92 (0.99) | 0.71 (0.87) | 0.45 (0.72) |
| Partner education | 11.81 (2.68) | 11.82 (2.89) | 11.71 (2.77) | 11.54 (2.95) |
| Partner race | | | | |
| White | 23.5% | 29.7% | 28.9% | 29.4% |
| Black | 34.9% | 25.3% | 22.3% | 17.4% |
| Hispanic | 39.0% | 41.0% | 45.8% | 51.6% |
| Relationship length | 9.63 (7.09) | 8.95 (6.43) | 9.85 (7.21) | 10.43 (8.10) |
| Number of children | 2.00 (1.14) | 2.09 (1.20) | 2.04 (1.20) | 2.16 (1.19) |
| Remarried | 27.1% | 21.1% | 17.5% | 15.0% |
| Race | | | | |
| White | 25.1% | 31.5% | 28.8% | 28.5% |
| Black | 34.6% | 25.9% | 22.3% | 17.3% |
| Hispanic | 37.0% | 39.9% | 46.3% | 51.0% |
| Partner relationship in trouble | 98.1% | 94.1% | 92.6% | 75.0% |
| Female | 55.4% | 52.7% | 54.1% | 44.5% |
| ATE[a] | 0.18 (.01) | 0.15 (.01) | 0.14 (.01) | 0.11 (.01) |
| Effect size[b] ($d$) | 0.08 | 0.07 | 0.07 | 0.05 |

*Note.* $N$ = 9,351. Standard errors are presented in parentheses after the mean. ATE = average treatment effect.
[a] ATE calculated from causal forest. [b] Calculated using *SD* from the original report.



the top predictors of intervention outcomes. However, pretreatment relationship happiness was only the second-best predictor of relationship happiness and the fourth-best predictor of negative emotions and behaviors.

Instead, pretreatment psychological distress was the factor most predictive of both treatment outcomes, and other top predictors included adverse childhood experiences and perceived stress. These results indicate that *individual* experiences and characteristics play a major role in the extent to which couples can benefit from relationship interventions. Yet the importance of individual risk factors has often been overlooked in this literature, with few past studies examining individual-level traits or experiences as moderators of treatment outcomes (cf. Carlson et al., 2017; Cooper et al., 2024; Williamson et al., 2015). The fact that this data-driven approach identified important predictors of treatment outcomes that have previously been overlooked highlights the shortcomings of top-down, researcher-driven approaches to understanding mechanisms of treatment outcomes and the promise of machine learning methods for unearthing future insights into intervention heterogeneity.

These findings also contribute to our understanding of the cumulative risk model in the context of CRE (Rauer et al., 2008). Results are consistent with the theory behind the cumulative risk model: Treatment outcomes must be predicted with a combination of risk factors and cannot be accurately predicted with only a single risk factor. However, the common operationalization of the cumulative risk model, which assumes that each risk factor conveys an equivalent impact on outcomes, was not supported. Instead, our findings indicated that some risk factors contributed disproportionately to heterogeneity in treatment effects, challenging the assumption of equifinality inherent to traditional cumulative risk indices. Instead, our results support differential weighting of risk factors, with personal characteristics playing a strong role. Additionally, our results were not consistent with the selection of risk factors that are typically included in risk indices. Multiple past studies have used a risk index comprised only of sociodemographic characteristics, such as age, income, and education (e.g., Amato, 2014; Ritchie et al., 2023; Williamson et al., 2016). However, the only demographic characteristic that was identified as important in the present study was education, which was the sixth most important predictor for negative emotions and behaviors.

Notably, although the causal forest analysis uncovered a range of treatment effects, the largest estimated treatment effects were still small in magnitude. In the original evaluation of the SHM program, the treatment effect for relationship happiness was $d$ = .13 (Lundquist et al., 2014), whereas the causal forest model estimated that the participants who benefited the most (i.e., the top quartile) had a treatment effect of $d$ = .28. However, it is worth noting that both studies focused on samples of low-income couples, a population in which treatment effects are typically very small. A meta-analysis of CRE delivered to low-income couples found a treatment effect of $d$ = .114 for relationship quality (Hawkins et al., 2022). Thus, the high-end effects observed in the SHM data are much larger than the typical outcome for similar participants and are closer to the ATE for middle-class couples ($d$ = .30; Hawkins et al., 2008).

**Table 7**
*Means and Standard Deviations of Quantile Groupings for Negative Emotions and Behaviors in Study 2*

| Pretreatment characteristic | Top 25% | 26%–50% | 51%–75% | Bottom 25% |
|---|---|---|---|---|
| Psychological distress | 2.46 (.62) | 1.83 (.50) | 1.43 (.31) | 1.16 (.19) |
| Partner psychological distress | 2.04 (.74) | 1.79 (.67) | 1.63 (.60) | 1.45 (.48) |
| Baseline relationship happiness | 6.61 (2.29) | 7.07 (1.93) | 8.0 (1.62) | 8.79 (1.25) |
| Partner relational aggression | 1.17 (1.39) | 0.62 (1.08) | 0.32 (.79) | 0.15 (.48) |
| Education | 12.41 (2.83) | 12.04 (2.77) | 11.96 (2.87) | 10.69 (2.56) |
| Number of children | 2.26 (1.21) | 2.09 (1.20) | 2.05 (1.20) | 1.92 (1.10) |
| Partner stressful events | 1.10 (1.08) | 0.76 (.92) | 0.81 (.95) | 0.54 (.75) |
| Partner income | 23,592 (26,760) | 21,804 (27,876) | 19,992 (17,748) | 17,592 (14,856) |
| Relational aggression | 1.01 (1.31) | 0.66 (1.10) | 0.39 (.87) | 0.17 (.56) |
| Income | 20,988 (30,336) | 22,152 (24,072) | 20,964 (18,528) | 18,612 (15,708) |
| Stressful events | 1.14 (1.02) | 0.87 (.97) | 0.73 (.91) | 0.48 (.75) |
| Partner baseline relationship happiness | 6.12 (2.33) | 7.06 (2.13) | 7.96 (1.79) | 8.41 (1.46) |
| Partner education | 11.80 (2.72) | 11.85 (2.81) | 11.83 (2.89) | 11.41 (2.86) |
| Relationship length | 9.76 (6.99) | 9.37 (6.91) | 10.2 (7.34) | 9.81 (7.89) |
| Relationship in trouble | 94.8% | 80.0% | 59.0% | 27.0% |
| Remarried | 24.7% | 22.9% | 17.9% | 14.6% |
| Partner race | | | | |
| White | 23.9% | 25.1% | 30.3% | 33.2% |
| Black | 31.1% | 26.8% | 22.5% | 19.0% |
| Hispanic | 43.1% | 44.6% | 43.8% | 45.6% |
| Race | | | | |
| White | 23.8% | 28.0% | 30.7% | 31.9% |
| Black | 31.1% | 27.4% | 22.1% | 18.8% |
| Hispanic | 42.0% | 42.5% | 44.3% | 45.8% |
| Female | 59.2% | 53.4% | 49.3% | 44.8% |
| Partner relationship in trouble | 97.5% | 95.4% | 85.6% | 74.0% |
| ATE[a] | −0.09 (.01) | −0.08 (.01) | −0.07 (.01) | −0.06 (.01) |
| Effect size[b] ($d$) | 0.13 | 0.11 | 0.10 | 0.08 |

*Note*. $N$ = 9,977. Standard errors are presented in parentheses after the mean. ATE = average treatment effect.
[a] ATE calculated from causal forest. [b] Calculated using *SD* from the original report.

There are several limitations that must be considered in the interpretation of this study. First, as mentioned above, the two samples used in these analyses were comprised of low-income couples. These data sets were chosen because the sample sizes were very large, which is required for stable estimation of CATEs (Wager & Athey, 2018), and the current funding landscape for relationship interventions means that the only large-scale studies of CRE being conducted are focused on low-income couples. Although this limits the generalizability of the results to higher socioeconomic status groups, it also enhances their relevance to the settings where the vast majority of CRE research and implementation are currently taking place.

Second, the use of an ITT approach uses all participants assigned to the treatment group, regardless of the dosage of the intervention they received. This has the advantage of providing a more conservative estimation of treatment effects. However, when examining questions about which subgroups benefit more, the ITT approach may have limitations. For example, it might overlook the fact that important participant characteristics could influence not only the treatment effect but also the likelihood of full participation in the intervention. For example, participant characteristics and program setting are associated with attendance rates of CRE interventions (Friend et al., 2023). These interactions between participant characteristics, intervention adherence, and treatment outcomes could potentially mask or distort subgroup-specific effects. Future research should consider each aspect (e.g., implementation, recruitment, attrition) with caution, especially when drawing applied conclusions from the results.

Third, there was variation in how constructs were measured and operationalized across the two studies. This inconsistency in assessment methods across programs restricted our ability to conduct uniform analyses across both interventions. Specifically, we were unable to examine the influence of an important characteristic, perceived stress, on treatment effects in PACT. However, this limitation is reflective of real-world implementation settings in which practitioners may have incomplete data or different measures of a construct but still wish to optimally allocate interventions. The fact that the trained models functioned well in the face of these limitations suggests that they were fairly robust.

Fourth, causal forest has emerged as a prominent method in the burgeoning field of machine learning–based causal inference, demonstrating excellent performance in estimating treatment effects on simulated data where traditional metrics (e.g., mean squared error or accuracy) can be evaluated. Nonetheless, a significant limitation of this approach is the lack of readily testable metrics for validating results on new, unseen data (Gong et al., 2021; Shiba & Inoue, 2024). We have utilized the currently available techniques to make inferences about the generalizability of the model trained in Study 1 to novel data in Study 2. However, future methodological development in this area would allow for new innovations, such as building a dashboard that CRE providers could use at intake to determine whether a couple is likely to benefit from their program. A version of this idea has been implemented in an online relationship intervention, which used random forests to develop an algorithm that predicts the level of coach support couples need while completing the program (Hatch & Doss, 2022).

Last, the observed outcomes may be influenced by ceiling effects. In truly preventive interventions, measurable benefits might not emerge within a 12-month timeframe unless the control group experiences substantial declines during that period. As such, these analyses may primarily detect benefits among couples who entered the program with lower functioning and thus have greater room for improvement. This focus on short-term outcomes may also obscure the potential long-term preventive benefits of the intervention, which could take years to manifest. This short-term lens makes it difficult to assess the preventive potential of relationship education programs, which were originally designed to maintain functioning and prevent future distress rather than produce immediate change. From a policy and funding perspective, this underscores the importance of supporting long-term follow-up in program evaluations. Without extended data collection, the full impact of preventive interventions cannot be accurately assessed.

In sum, this research contributes to the growing body of evidence supporting the examination of heterogeneity in treatment effects of social and psychological interventions and demonstrates a cutting-edge machine learning approach for doing so (Bolger et al., 2019; Bryan et al., 2021). These studies provide evidence that the impacts of CRE interventions are heterogeneous, and treatment outcomes can be predicted by pretreatment characteristics. CRE interventions yield greater benefits for individuals experiencing higher levels of individual and relational distress when they enter the program, whereas those with better baseline functioning benefit very little from CRE. By elucidating the complex interplay of pretreatment factors and their impact on intervention outcomes, this research advances our theoretical understanding of CRE effectiveness and points toward future directions to improve the dissemination and implementation of this type of intervention.

## References


Amato, P. R. (2014). Does social and economic disadvantage moderate the effects of relationship education on unwed couples? An analysis of data from the 15-month Building Strong Families evaluation. *Family Relations*, *63*(3), 343–355. https://doi.org/10.1111/fare.12069

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7353–7360. https://doi.org/10.1073/pnas.1510489113

Bolger, N., Zee, K. S., Rossignac-Milon, M., & Hassin, R. R. (2019). Causal processes in psychology are heterogeneous. *Journal of Experimental Psychology: General*, *148*(4), 601–618. https://doi.org/10.1037/xge0000558

Bradford, A. B., Drean, L., Adler-Baeder, F., Ketring, S. A., & Smith, T. A. (2017). It's about time! Examining received dosage and program duration as predictors of change among non-distressed and distressed married couple and relationship education participants. *Journal of Marital and Family Therapy*, *43*(3), 391–409. https://doi.org/10.1111/jmft.12209

Braithwaite, S., & Holt-Lunstad, J. (2017). Romantic relationships and mental health. *Current Opinion in Psychology*, *13*, 120–125. https://doi.org/10.1016/j.copsyc.2016.04.001

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, *300*, 70–79. https://doi.org/10.1016/j.neucom.2017.11.077

Carlson, R. G., Rappleyea, D. L., Daire, A. P., Harris, S. M., & Liu, X. (2017). The effectiveness of couple and individual relationship education: Distress as a moderator. *Family Process*, *56*(1), 91–104. https://doi.org/10.1111/famp.12172

Cherlin, A. (2010). Demographic trends in the United States: A review of research in the 2000s. *Journal of Marriage and Family*, *72*(3), 403–419. https://doi.org/10.1111/j.1741-3737.2010.00710.x

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India* (No. w24678). National Bureau of Economic Research. https://doi.org/10.3386/w24678

Cooper, E., Adler-Baeder, F., & McGill, J. (2024). Adverse childhood experiences among couple relationship education participants: Exploring influence on program impact and testing models of risk and resilience. *Journal of Couple & Relationship Therapy*, *23*(4), 423–446. https://doi.org/10.1080/15332691.2024.2345686

Don, B. P., Simpson, J. A., Fredrickson, B. L., & Algoe, S. B. (2025). Interparental positivity spillover theory: How parents' positive relational interactions influence children. *Perspectives on Psychological Science*, *20*(4), 639–668. https://doi.org/10.1177/17456916231220626

Doss, B. D., Cicila, L. N., Georgia, E. J., Roddy, M. K., Nowlan, K. M., Benson, L. A., & Christensen, A. (2016). A randomized controlled trial of the web-based OurRelationship program: Effects on relationship and individual functioning. *Journal of Consulting and Clinical Psychology*, *84*(4), 285–296. https://doi.org/10.1037/ccp0000063

Fletcher, G. J. O., Simpson, J. A., Campbell, L., & Overall, N. C. (2015). Pair-bonding, romantic love, and evolution: The curious case of Homo sapiens. *Perspectives on Psychological Science*, *10*(1), 20–36. https://doi.org/10.1177/1745691614561683

Friend, D., Cornwell, D., Marks-Anglin, A., Hennigar, A., Troxel, J., & Roemer, G. (2023). *Predicting participation in healthy marriage and responsible fatherhood programs* (No. OPRE Report #2023-125). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Gong, X., Hu, M., Basu, M., & Zhao, L. (2021). Heterogeneous treatment effect analysis based on machine-learning methodology. *CPT: Pharmacometrics & Systems Pharmacology*, *10*(11), 1433–1443. https://doi.org/10.1002/psp4.12715

Halford, W. K., & Bodenmann, G. (2013). Effects of relationship education on maintenance of couple relationship satisfaction. *Clinical Psychology Review*, *33*(4), 512–525. https://doi.org/10.1016/j.cpr.2013.02.001

Halford, W. K., Markman, H. J., & Stanley, S. (2008). Strengthening couples' relationships with education: Social policy and public health perspectives. *Journal of Family Psychology*, *22*(4), 497–505. https://doi.org/10.1037/a0012789

Halford, W. K., Rahimullah, R. H., Wilson, K. L., Occhipinti, S., Busby, D. M., & Larson, J. (2017). Four year effects of couple relationship education on low and high satisfaction couples: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, *85*(5), 495–507. https://doi.org/10.1037/ccp0000181

Hatch, S. G., & Doss, B. D. (2022). *Optimizing the amount and type of practitioner contact in online relationship education for couples with low incomes: Deploying machine learning models to the web* (Behavioral Interventions Scholars). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hawkins, A. J., Blanchard, V. L., Baldwin, S. A., & Fawcett, E. B. (2008). Does marriage and relationship education work? A meta-analytic study. *Journal of Consulting and Clinical Psychology*, *76*(5), 723–734. https://doi.org/10.1037/a0012584

Hawkins, A. J., Hokanson, S., Loveridge, E., Milius, E., Duncan, M., Booth, M., & Pollard, B. (2022). How effective are ACF-funded couple relationship education programs? A meta-analytic study. *Family Process*, *61*(3), 970–985. https://doi.org/10.1111/famp.12739

Hsueh, J., Alderson, D. P., Lundquist, E., Michalopoulos, C., Gubits, D., Fein, D., & Knox, V. (2012). *The Supporting Healthy Marriage evaluation: Early impacts on low-income families* (No. OPRE Report #2012-11). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hsueh, J., & Knox, V. (2014). *Supporting healthy marriage evaluation: Eight sites within the United States, 2003–2013. Inter-university consortium for political and social research* [distributor]. https://doi.org/10.3886/ICPSR34420.v2

Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin*, *118*(1), 3–34. https://doi.org/10.1037/0033-2909.118.1.3

Kurz, C. F. (2022). Augmented inverse probability weighting and the double robustness property. *Medical Decision Making*, *42*(2), 156–167. https://doi.org/10.1177/0272989X211027181

Lundquist, E., Hsueh, J., Lowenstein, A. E., Faucetta, K., Gubits, D., Michalopoulos, C., & Knox, V. (2014). *A family strengthening program for low-income families: Final impacts from the supporting healthy marriage evaluation* (No. OPRE Report #2014-09A). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Maisel, N. C., & Karney, B. R. (2012). Socioeconomic status moderates associations among stressful events, mental health, and relationship satisfaction. *Journal of Family Psychology*, *26*(4), 654–660. https://doi.org/10.1037/a0028901

McNulty, J. K., Meltzer, A. L., Neff, L. A., & Karney, B. R. (2021). How both partners' individual differences, stress, and behavior predict change in relationship satisfaction: Extending the VSA model. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(27), Article e2101402118. https://doi.org/10.1073/pnas.2101402118

Miller Gaubert, J., Gubits, D., Principe Alderson, D., & Knox, V. (2012). *The supporting healthy marriage evaluation: Final implementation findings* (No. OPRE Report #2012-12). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Moore, Q., Avellar, S., Patnaik, A., Covington, R., & Wu, A. (2018). *Parents and children together: Effects of two healthy marriage programs for low-income couples* (No. OPRE Report #2018-58). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Quintana, R. (2023). Embracing complexity in social science research. *Quality & Quantity: International Journal of Methodology*, *57*(1), 15–38. https://doi.org/10.1007/s11135-022-01349-1

Quirk, K., Strokoff, J., Owen, J. J., France, T., & Bergen, C. (2014). Relationship education in community settings: Effectiveness with distressed and non-distressed low-income racial minority couples. *Journal of Marital and Family Therapy*, *40*(4), 442–453. https://doi.org/10.1111/jmft.12080

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing.

Rauer, A. J., Karney, B. R., Garvan, C. W., & Hou, W. (2008). Relationship risks in context: A cumulative risk approach to understanding relationship satisfaction. *Journal of Marriage and Family*, *70*(5), 1122–1135. https://doi.org/10.1111/j.1741-3737.2008.00554.x

Ritchie, L. L., Stanley, S. M., Allen, M. O. T., & Rhoades, G. K. (2023). Impact evaluation of the family expectations program and moderation by sociodemographic disadvantage. *Family Process*, *62*(1), 201–215. https://doi.org/10.1111/famp.12762

Robles, T. F., Slatcher, R. B., Trombello, J. M., & McGinn, M. M. (2014). Marital quality and health: A meta-analytic review. *Psychological Bulletin*, *140*(1), 140–187. https://doi.org/10.1037/a0031859

Rogge, R. D., Cobb, R. J., Lawrence, E., Johnson, M. D., & Bradbury, T. N. (2013). Is skills training necessary for the primary prevention of marital distress and dissolution? A 3-year experimental study of three interventions. *Journal of Consulting and Clinical Psychology*, *81*(6), 949–961. https://doi.org/10.1037/a0034209

Shiba, K., & Inoue, K. (2024). Harnessing causal forests for epidemiologic research: Key considerations. *American Journal of Epidemiology*, *193*(6), 813–818. https://doi.org/10.1093/aje/kwae003

Tibshirani, J., Athey, S., Sverdrup, E., & Wager, S. (2022). *Generalized random forests* [Computer software]. https://grf-labs.github.io/grf/

Urganci, B., Chen, P.-H., & Williamson, H. C. (2024). Many distressed couples seek community-based relationship education but don't benefit from it. *Family Relations*, *73*(4), 2510–2526. https://doi.org/10.1111/fare.13015

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Williamson, H. C., Altman, N., Hsueh, J., & Bradbury, T. N. (2016). Effects of relationship education on couple communication and satisfaction: A randomized controlled trial with low-income couples. *Journal of Consulting and Clinical Psychology*, *84*(2), 156–166. https://doi.org/10.1037/ccp0000056

Williamson, H. C., Rogge, R. D., Cobb, R. J., Johnson, M. D., Lawrence, E., & Bradbury, T. N. (2015). Risk moderates the outcome of relationship education: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *83*(3), 617–629. https://doi.org/10.1037/a0038621

Zaveri, H., & Baumgartner, S. (2016). *Parents and children together: Design and implementation of two healthy marriage programs* (No. OPRE Report #2016-63). Mathematica Policy Research.

Received July 25, 2025
Revision received October 15, 2025
Accepted October 17, 2025 ■