EVALUATION OF SECONDARY PREVENTION IN THE COMMUNITY, FAMILY AND YOUTH RESILIENCE (CFYR) PROGRAM IN ST. LUCIA, ST. KITTS AND NEVIS AND GUYANA

Final Endline Report

August 2020

This publication was prepared independently by Alberto Díaz-Cayeros, Stephanie Gimenez Stahlberg, Rachel Pizatella-Haswell, Daniel Sabet, and Julia Kresky of Social Impact. It was produced at the request of the United States Agency for International Development as part of the Democracy, Human Rights, and Governance – Learning, Evaluation, and Research activity.

EVALUATION OF SECONDARY PREVENTION IN THE COMMUNITY, FAMILY AND YOUTH RESILIENCE (CFYR) PROGRAM IN ST. LUCIA, ST. KITTS AND NEVIS AND GUYANA

Final Endline Report

August 2020

AID-OAA-M-13-00011

Suggested Citation:

Diaz-Cayeros, Alberto, Stephanie Gimenez Stahlberg, Rachel Pizatella-Haswell, Daniel Sabet, and Julia Kresky (2020, August). Evaluation of Secondary Prevention in the Community, Family and Youth Resilience (CFYR) Program in St. Lucia, St. Kitts and Nevis and Guyana. Arlington, VA: Social Impact, Inc.

ACKNOWLEDGEMENTS

This evaluation has required extensive coordination between many different groups and individuals. First and foremost, we would like to thank our colleagues at the United States Agency for International Development (USAID) for continually pushing and advocating for a rigorous evaluation design despite the many challenges this entailed. We are grateful to Ryssa Brathwaite, Mansfield Blackwood, Andrew Greer, Morgan Holmes, Ted Lawrence, Chloe Noble, Sharon Ramsaran, Angela Skeete and Brandy Witthoft. The evaluation required the implementer Creative Associates and Community, Family, and Youth Resilience (CFYR) to invest a great deal of time and energy into ensuring strong programming while adapting to put the right conditions in place for a rigorous evaluation. We want to thank Debra Wahlberg, Vaughn Graham, K. Angus McMillan, Bert Laurent, Dwynette Eversley, Dorn Henry, Courtney Brown, Desmona Jackson, Erik Alda, Alexandra Simonians, Jason Wilks, Carlo Arze, Donald Cole, Waqas Mahmood, Sara Danish, Enrique Roig, Guillermo Cespedes, all of the *Family Matters* family counselors, Chuck Katz of Arizona State University, and University of Southern California (USC) colleagues Karen Hennigan and Kathy Kolnick. The National Opinion Research Center (NORC) at the University of Chicago led much of the data collection, and we appreciate the efforts of Ron Wendt, Kareem Kysia, Alexander Rigaux, Xiran Liu, Letitia Onyango, Renee Hendley and Jeff Telgarsky. We would also like to thank data collection partners in the three countries, including The Solomon Group, Specialists in Sustained Youth Development and Research (SSYDR) and the St. Lucia Central Statistical Office, as well as Social Impact (SI) field coordinators Carol English, Terry Morris, Lydia Osbourne, Sacha Harris, and Marecia Pemberton, qualitative fieldwork consultants Wemyss de Florimonte, Zahra Jacobs and Andrew Grant, and former SI staff who supported the Democracy, Human Rights, and Governance – Learning, Evaluation and Research (DRG-LER) mechanism, Jordan Robinson, Alison Miranda and Meredith Feenstra. Finally, we are grateful to the many youth, parents, and caregivers who participated in the Youth Services Eligibility Tool (YSET) assessment and caregiver survey.

CONTENTS

TABLES

FIGURES

ACRONYMS

Acronym	Definition
CARICOM	Caribbean Community
CBT	Cognitive Behavioral Therapy
CCYD	CARICOM Commission on Youth Development
CEM	Coarsened Exact Matching
CFYR	Community, Family and Youth Resilience
DRG-LER	Democracy, Human Rights, and Governance – Learning, Evaluation, and Research
EGAP	Evidence in Governance and Politics
FFT	Functional Family Therapy
GLS	Generalized Least Squares
GRYD	Gang Reduction and Youth Development
GUY	Guyana
IDB	Inter-American Development Bank
ITT	Intention-to-treat
LAC	Latin American and Caribbean
LATE	Local Average Treatment Effect
LST	LifeSkills Training
MDE	Minimum Detectable Effect
MFD	Model Fidelity Database
MST	Multisystemic Therapy
NORC	National Opinion Research Center
OLS	Ordinary Least Squares
PAP	Poverty Alleviation Program
PIFSM	Prevention and Intervention Family Systems Model
RCT	Randomized Controlled Trial
SI	Social Impact
SKN	St. Kitts and Nevis
SLU	St. Lucia
SSYDR	Specialists in Sustained Youth Development and Research
TOT	Treatment-on-the-treated
TYG	Troublesome Youth Group
USAID	United States Agency for International Development
USC	University of Southern California
WHO	World Health Organization
YDP	YMCA Youth Development Program
YSET	Youth Services Eligibility Tool

EXECUTIVE SUMMARY

Social Impact (SI) was contracted by the United States Agency for International Development (USAID) to evaluate the Community, Family and Youth Resilience (CFYR) program in Guyana (GUY), St. Kitts and Nevis (SKN), and St. Lucia (SLU). The initiative focuses on five communities with relatively high levels of crime and violence in each country for a total of 15 communities. Across these 15 communities, the CFYR implementer Creative Associates employs a public-health informed, place-based strategy to reduce and prevent youth risky, antisocial and violent behavior and attitudes. CFYR programming includes primary prevention, focused on the general youth population, secondary prevention, focused on youth at a higher risk level, and tertiary prevention, aimed at those who have already been "infected" by crime and violence. This impact evaluation focuses on CFYR's secondary prevention efforts, the Family Matters program.

USAID/CFYR's *Family Matters* intervention (*Family Matters*) is a secondary violence prevention method that targets youth between 10 and 17 years of age who are at a secondary or tertiary level of risk of engaging in delinquent behaviors leading to crime and violence. Eligible youth are identified through the Youth Services Eligibility Tool (YSET) assessment, which is used to determine their level of risk. *Family Matters* engages beneficiary youth and their families in an approximately seven-month cycle of structured family counseling, specifically adapted for the Caribbean context. There are two cycles, for a total of 12-14 months of intervention.

The intervention is grounded in research that shows that positive behavioral changes in youth are more likely to last if embedded in and reinforced by the family and the larger community. Throughout the implementation cycle, family counselors trained by USAID/CFYR in the *Family Matters* methodology hold regular face-to-face counseling sessions with participating youth and their families. During these meetings, typically held in the family's home, counselors strengthen family cohesion (bonds that hold family members together) and connect the family to the wider community. Family counselors also work with youth on an individual basis to help them adopt positive and safe behavior. The intervention being evaluated in this report was implemented between June 2018 and December 2019 in the three countries.

CFYR identified higher risk youth for the program using the YSET, a tool originally designed by researchers at the University of Southern California (USC) to gauge the risk of youth joining gangs in Los Angeles and later adapted to the Caribbean context. The survey includes behavioral and attitudinal scales, nine of which are used for calculating program eligibility. CFYR and its local partners, along with SI and the National Opinion Research Center at the University of Chicago (NORC), worked to identify the population of youth in the targeted communities and apply the YSET tool. Program eligibility was calculated by CFYR and USC. Eligible youth, those at higher risk, were then randomized into two groups, one that would receive the Family Matters intervention, and a second group that would serve as a control and be eligible to receive the intervention the following year. As such, this evaluation is a multi-site randomized controlled trial (RCT) where sites were predetermined but youth were randomized within each site.

To assess impact, the evaluation compares two data sources and diverse outcome measurements over time and across the treatment and control groups. Youth in the treatment were assigned to receive the intervention at this time (and may or may not have completed it), and control youth did not receive the intervention. The first data source is the YSET itself, which was reapplied with eligible treatment and control youth after approximately six months of the intervention and at completion of the program after 12 months. Specifically, the evaluation uses the YSET to examine potential impacts across nine modules

measuring youth risk and make an overall risk determination. The second data source is a survey of the youths' primary caregivers, which was implemented prior to the intervention and again after completion. The caregiver survey allows us to examine impact from the perspective of caregivers across several modules related to the intervention. The team created the survey by drawing on relevant YSET modules and producing additional questions and modules that match expected outcomes from the program.

We also draw on a community survey implemented by CFYR and its partners in all 15 communities and a few comparison communities. While we do not use this survey to assess impact, it provides valuable contextual information across communities about victimization, perceptions of crime prevalence and social capital. We use this survey to check for community-wide changes over the course of the program, which would have affected both treatment and control youth and their families.

At baseline, 5,109 youth were assessed with the YSET, and 1,113 were determined to be eligible for the program. The eligible youth were randomized into treatment (566 youth) and control (567) groups. At endline, NORC and its partners re-assessed 499 youth assigned to the control group and 481 youth assigned to treatment, out of which 230 (48 percent) had completed the intervention. The remaining youth had either dropped out (20 percent), declined the intervention (16 percent), or not been invited to participate (16 percent).

The caregiver sample is composed of mothers, fathers, guardians and other relatives or non-relatives responsible for the youth. Unlike youth, caregivers were only surveyed at baseline and endline. The total caregiver sample at baseline was 1,002, with 501 control caregivers and 501 treatment caregivers. Enumerators surveyed 1,024 caregivers at endline, which was more than were surveyed at baseline because there were fewer refusals. Of the caregivers who were surveyed at endline, 240 of them had youth who completed the program. The YSET and caregiver analysis both use panel data, meaning that the samples only include the respondents that were interviewed both at baseline and endline. We use a difference-in-differences approach and ran the analysis on a matched sample when the sample was determined to be unbalanced.

Older (ages 15 and above) and out-of-school youth were more likely to drop out of the program (by nine and 24 percent, respectively). This finding may be concerning considering that out-of-school youth are generally at a higher risk.[1] We did not find other differences in attrition across gender, overall risk level or risk level for a particular YSET scale. Youth who scored at risk for the self-reported delinquency scale (IJ) were less likely to decline participation than those who did not score at risk on this module.

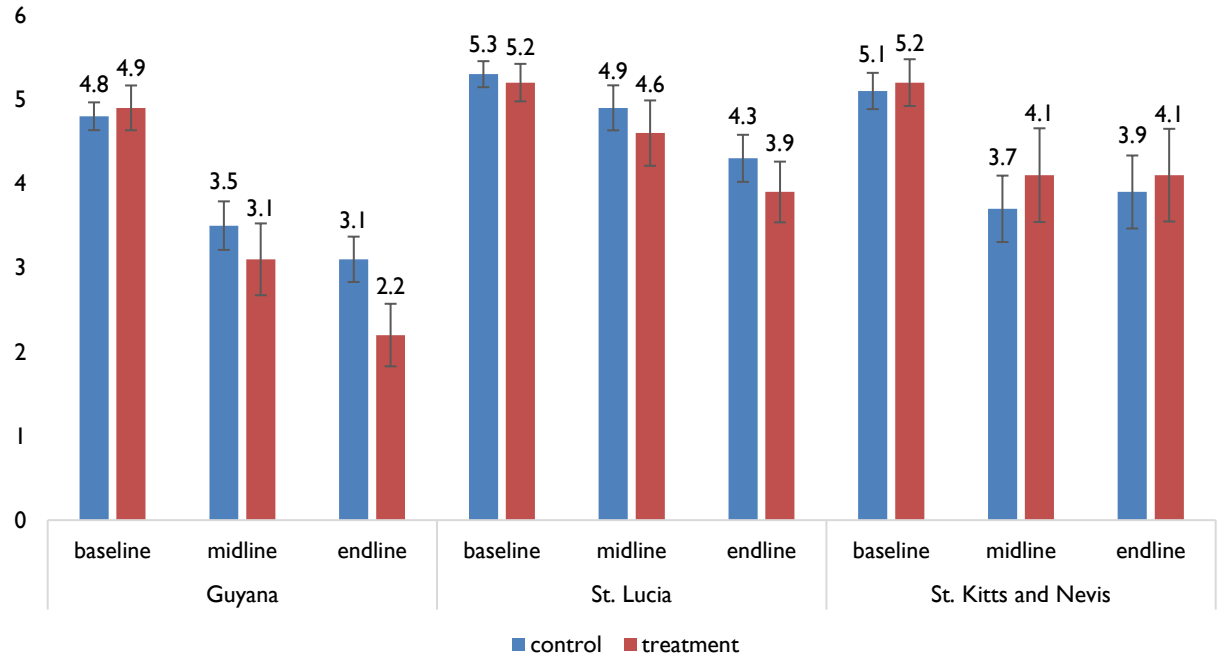
Both treatment and control youth experience a substantial drop in estimated risk levels between baseline and endline. There are, however, some statistically significant differences between treatment and control youth from baseline to endline. On average, treatment youth experienced a 34 percent reduction in their mean number of risk factors from baseline to endline (from 5.12 to 3.37), and control youth experienced a 26 percent reduction (from 5.09 to 3.75). In other words, treatment youth experienced a reduction that is 0.42 points larger (in the zero to nine YSET scale) than that of the control group. In addition, the reduction in the share of treatment youth who are "eligible" (four or more risk factors) was 9.3 percent larger than the reduction for the control group. There are statistically significant results across some of

[1] Among baseline respondents, the mean number of risk factors for youth who reported being in school was 1.98, and the mean score for youth who reported being out of school was 3.16. This difference in means indicates that out-of-school youth are generally at a higher risk.

the YSET scales as well. Although, scale F measuring rationalization of behavior yielded the only relatively robust results where the difference between treatment and control is statistically significant both in terms of average score changes and the share of youth scoring at risk on that scale. In general, most of the individual scales showed no statistically significant difference between youth in the program and youth who were not in the program.

The statistically significant results in the "mean number of risk factors" and "eligible" outcomes are primarily driven by Guyana. Figure 1 shows that the differences between treatment and control at endline are quite similar for St. Lucia and St. Kitts and Nevis but are larger in Guyana. In St. Kitts and Nevis, the average scores of treatment youth increased slightly from baseline to endline. In Guyana, 84 percent of treatment youth saw a decrease in their number of risk factors from baseline to endline and only seven percent saw an increase in their number of risk factors from baseline to endline. By comparison, in St. Kitts and Nevis, 29 percent treatment youth saw an increase in their risk factors from baseline to endline.

Figure 1: Average number of risk factors by data collection period for TOT sample, by country (with standard error bars)



We estimate the main effects of the program using four different models. All of these models when run on the full treatment-on-the-treated (TOT) sample, yield a coefficient close to 0.42 on the mean number of risk factors scale. The effect is small, considering there are nine scales that make up this variable, an eligible youth scored at least four on this scale at baseline, and youth in both treatment and control groups saw an average change of about two points on this scale from baseline to endline. The effect is also small, considering the standard deviation of the YSET scale is 2.12. Notably, an alternative estimation approach that matches a smaller control group to the treatment group in order to improve balance finds an even smaller and not statistically significant relationship between the program and youth's mean number of risk factors.

To examine the impact of the *Family Matters* program on caregivers, we look at indicators that capture family dynamics and the relationship between the caregiver and youth through the caregiver survey. The survey

measures outcomes expected to result from the *Family Matters* program and was validated for the Caribbean context. Empirically, the program does not result in caregiver improvements across the outcomes of interest. Changes that treatment caregivers experience from baseline to endline are not statistically different from the changes that control caregivers experience. However, treatment caregivers overwhelmingly review *Family Matters* and the counselors favorably, which suggests that caregivers gained some benefit from the program even if it is not captured empirically. Youth participants also evaluate the intervention positively, although respondents from St. Kitts and Nevis are less likely than youth from the other countries to recommend the *Family Matters* program to others.

In summary, over the time that the program was implemented, we see a strong decrease in risk among youth in all countries, regardless of whether they were in the treatment or the control group. It is only for a few indicators that the decrease in risk is slightly larger for treatment youth than for control youth. We acknowledge the positive assessments of the program by youth and caregivers, but empirically, we can only conclude with confidence that the intervention had a minor impact on youth risk factors. Our main results aggregate the samples from the three countries, as each country sample on its own is not large enough to reach conclusions with confidence. We checked that our analysis had enough power to detect small effects on each outcome by calculating minimum detectable effects (MDEs). We conclude that the outcomes with null findings are not a result of a lack of power, but rather the lack of a meaningful difference between treatment and control groups.

When discussing the findings, we explore potential reasons why the control group may have experienced such a large reduction in measured risk. We find no evidence of changes at the community level in trust, victimization, social capital or crime perceptions that may account for a shift in environment in communities that would influence youth risk, thus confounding our estimates. Based on a literature review, we find compelling evidence that youth may be more inclined to want to give "correct answers," or satisfice, particularly in such a long survey (the YSET used for the countries in this study had 17 modules). However, we do not believe these effects would be stronger on control youth. We cannot completely rule out some spillover effects given the proximity of treatment and control youth, but we do not find evidence of spillovers in our analyses of survey questions and focus group discussions. In addition, any spillovers are unlikely to influence control youth enough to attenuate the results. It is possible that both treatment and control youth shift behavior as a consequence of the added attention they receive when being observed, a phenomenon known as the Hawthorne effect. However, we do not believe such an effect can account for the large shift in the control group. The most likely explanation is that treatment and control youth's risk level is similar at endline, and the drop in risk for both groups concerns the way they relate to the YSET as a long instrument that gets repeated three times over the course of the program.

The impact evaluation teaches lessons that cannot be obtained from a simpler assessment of client satisfaction or from an observational study tracking youth through time without a control group comparison. Some of the recommendations we make are more likely to be relevant for governments. Other recommendations may be useful to USAID seeking to support similar programs in the region, and others may be important for agencies and partners doing implementation on the ground.

The evaluation has raised concerns with the YSET as both an evaluation and assessment tool. There is some evidence to suggest that youth were not engaged throughout the assessment, that the instrument is long, and that some of the modules are not clearly linked to the intervention. While it is reasonable to adapt a validated instrument from another context, the YSET was validated for a different purpose and a

different context. As such, we recommend further exploration of the YSET as an evaluation and assessment survey, and potentially developing a revised tool. The subsequent tool should be 1) shorter, 2) tested for survey fatigue, 3) tested for producing reliable results (including tests for respondent comprehension of questions), and 4) validated as a predictor of youth violence.

While the qualitative data suggests that the intervention was well implemented, we would imagine that future initiatives will incorporate learning from this initial intervention. As such, even though we find only a minor impact on potential youth violence prevention, we do not recommend terminating family counseling-based programming and instead recommend adjusting the intervention and looking for opportunities to re-evaluate future iterations.

INTRODUCTION

Social Impact (SI) was contracted by the United States Agency for International Development (USAID) to evaluate the Community, Family and Youth Resilience (CFYR)'s program in Guyana (GUY), St. Kitts and Nevis (SKN), and St. Lucia (SLU). The initiative focuses on five communities with relatively high levels of crime and violence in each country for a total of 15 communities.[2]

Crime and violence are growing problems in the Caribbean region. With increased globalization, the region has experienced a proliferation of trans-national crime and drug-trafficking that is unprecedented in scope and scale (CARICOM, p. 16). A Caribbean Community (CARICOM) Commission on Youth Development (CCYD) found that crime and violence are primary concerns for youth in the Caribbean, as young people speak of "fear, perceptions of lack of safety and concern for their general well-being as a result of the increased crime and violence; of self-imposed curfews, diminished participation in community activities, restriction of night-time activities, and changes in social practice as a consequence" (CARICOM, pp. xvi-xvii). Constant exposure to crime and violence can deeply affect a person's wellbeing, through increases in stress levels, emotional scarring, grief, and by creating a sense of loss (CARICOM, p. xvi).

Although the level of social and economic development varies across the three countries included in this project, all three have marginalized communities where youth face significant hardships at the personal, family and community levels. Each of the countries exhibits profound dislocations in their family structures, including a high prevalence of single-mother-headed households, missing male role models, and a relatively large transient population due to migration.[3] Some of these communities are afflicted by local gangs that control drug sales and instigate violence. The communities are all urban and some are low-income, although education, health and public services provisions are varied across these communities.

The main goal of the CFYR intervention is to reduce risky youth behaviors and attitudes, and ultimately youth violence. Implemented by Creative Associates, CFYR employs a public-health informed, place-based strategy. Consistent with the public health approach, CFYR programming includes primary prevention, focused on the general youth population, secondary prevention, focused on youth at a higher risk level, and tertiary prevention, aimed at offenders who have already been involved in crime and violence. This impact evaluation focuses on CFYR's secondary prevention efforts rather than CFYR as a whole.

The *Family Matters* secondary prevention program is targeted towards high-risk youth (those more likely to engage in criminal and violent activities), and it is one of four components of CFYR's programing. *Family Matters* consists of counseling sessions for youth aged 10-17 (at the start of the program) and their families. Family counselors who have been trained on the Prevention and Intervention Family Systems Model (PIFSM) engage with the families to strengthen family cohesion (bonds that hold family members together) and connect the family to the wider community. Family counselors also work with youth on an individual basis to help them adopt positive and safe behavior.

Enumerators visited schools identified by CFYR to educate students from treatment communities in order to identify youth eligible for the program. After gaining proper consent from their parents or guardians,

[2] A list of communities included in this intervention can be found in the "Background on the CFYR *Family Matters* Program" section.

[3] Census and other national data for three countries was analyzed as background for this evaluation.

enumerators interviewed youth using the Youth Services Eligibility Tool (YSET). The YSET was originally designed by researchers at the University of Southern California (USC) to gauge the risk of youth joining gangs in Los Angeles and was later adapted to the Caribbean context. CFYR and its local partners, along with SI and the National Opinion Research Center at the University of Chicago (NORC), worked to identify the population of youth in the targeted communities and apply the YSET tool. Program eligibility was calculated by CFYR and USC. Eligible youth at higher risk were then randomized by SI into two groups, one that would receive the *Family Matters* intervention and a second group that would serve as a control and be eligible to receive the intervention the following year. As such, this evaluation is a multi-site randomized controlled trial (RCT) where sites were predetermined but youth were randomized within each site.

The YSET and an additional caregiver survey were applied a few months before the intervention started. The intervention began in Guyana in June 2018, in St. Lucia in July 2018, and in St. Kitts and Nevis a month later, in August 2018. At those points in time, families whose youth were eligible for the program and who had been randomized into a treatment group for the purposes of the evaluation were approached, and the family counseling started soon after. A control group of youth who would be eligible to receive a delayed intervention was also identified. These youth were followed over the course of the evaluation.

Guyana, St. Lucia and St. Kitts and Nevis provide unique opportunities to explore programs designed to prevent youth engagement with criminal activity and to reduce youth violence. While we know much about how these processes unfold in cities and neighborhoods in the United States (and increasingly more knowledge is being generated in Central America), the opportunity to study prevention strategies in these communities may shed light on how to create successful programs that can impact local social dynamics in many other developing countries, including the rest of the Caribbean.

Two of the countries chosen for the intervention are very small, rendering it unfeasible to randomize treatment across communities. It would be close to impossible to find a large enough number of comparable control communities (in order to ensure power) and geographically far enough (to ensure no spillovers) that are similar enough to those chosen for the intervention. This has been done in large countries like India where there are literally tens of thousands, if not hundreds of thousands of potential communities. Thus, the evaluation team designed a multi-site RCT, whereby youth were randomized into control or treatment groups within the 15 pre-selected, high-risk communities across the three countries.[4] As such, the impact evaluation allows the program to encompass all of the selected communities, rather than a random selection of communities within hotspot territorial areas that would leave some selected communities untreated.

It is only natural that there is always pressure for a program evaluation to yield positive findings. Substantial effort is put into implementing a program, scarce national budget and international aid financial resources are spent, and there is a commitment by stakeholders to show that the effort and resources were not wasted. A randomized evaluation provides evidence of what can be scientifically ascertained regarding the program's effects on measurable outcome indicators. This means that there might be other outcomes that the study did not measure or nuances of the intervention that are not captured by average or mean effects. However, it is important that an evaluation study avoid yielding false positives, in the sense of reporting

[4] In Guyana, at the request of the Government of Guyana, the communities selected for the intervention were not necessarily the ones at highest risk, as CFYR excluded from its selection process several high-risk communities with programming implemented by the Inter-American Development Bank (IDB). In St. Lucia, due to small population sizes and limited disaggregated data to inform community selection, some of the communities include areas from different parishes/counties.

findings that are only statistically significant under unique conditions determined by the methodology. A scientific study must also ensure enough statistical power in its design to avoid false negatives, in the sense of being unable to measure effects that may exist but may be missed due to the size of the sample. A substantial amount of care has been put in this evaluation to ensure we do not make either of those mistakes. To avoid false positive results, we run the analysis on different samples (ITT, TOT, and a matched sample), and use different models to compare results. To avoid false negative results, we check that we have enough power for all of the main outcomes by calculating ex-post minimum detectable effects (MDEs).

This endline report begins with a brief literature review on the state of youth violence and violence prevention policies and programs. Next, we describe the evaluation purpose and evaluation questions, background on the CFYR program, evaluation design, methods and limitations, and, finally, the data analysis findings. The results include findings on youth behavior and attitudes from YSET analysis, as well as information on the home environment and parenting style of caregivers from the caregiver survey, comparing the results from baseline to endline. In addition, we also report youth and caregiver assessments of the *Family Matters* program, perform analysis of targeting and attrition, discuss the possibility of spillover effects, and include a discussion on the reasons why the control group may show improvements over time even when it does not receive the intervention.

LITERATURE REVIEW ON REDUCING YOUTH RISK LEVELS FOR ENGAGING IN CRIME AND VIOLENCE

THE STATE OF YOUTH VIOLENCE IN THE CARIBBEAN

Citizen security has become a greater concern for residents of the Caribbean, where the level of violent crime is particularly high and on the rise. Violent crime, including homicides, assaults and threats, appears to be more common in the Caribbean than in all other regions covered by the International Crime Victimization Survey database. Nearly one in three Caribbean residents reported witnessing a violent act in which another person got injured or died. Youth (18-24) and young adults (25-30) are disproportionally likely to be victims or perpetrators of violent crime, when one considers their share of the population overall (Sutton and Ruprah, 2017).

A 2013 Caribbean youth assessment prepared by SI revealed respondents' perception of heightened crime and violence involving youth in the region. Thefts and house break-ins, which are considered to be the predominant youth crimes, were on the rise and becoming more sophisticated as of 2013. Key informants in St. Vincent and the Grenadines, Dominica and St. Lucia reported that thieves hand off stolen articles to collaborators on boats who transport and sell them in other islands. This level of organization within local and inter-island networks also suggests that a gang problem is on the rise (Foss et al., 2013).

Perception of high levels of youth involvement in crime and violence in the Caribbean appears to be based on fact. A 2016 report titled "Prevalence and Patterns of Troublesome Youth Groups in the Caribbean" found that youth themselves report high levels of involvement in crime and violence. Two-thirds of surveyed students (across nine Caribbean countries[5]) reported having committed criminal and/or violent offenses in the previous year, and this number was even higher for students who acknowledged being gang members in Guyana and St. Lucia. For instance, involvement in violence was seven times higher in Guyana among gang members, and five times higher in St. Lucia (Katz and Nuño, 2016). This report suggests that gang involvement starts at a very young age, as early as nine years old.

THE CAUSES OF YOUTH VIOLENCE IN THE REGION

Set against a backdrop of drug trafficking, organized crime, weak judicial and law enforcement systems, and a lack of opportunities and support for youth who live in deprived communities, Laura Chioda (2017), Foss et al. (2013), and Katz and Nuño (2017) identify several factors that drive youth participation in crime and violence in the Latin American and Caribbean (LAC) region. Laura Chioda (2017) finds that in LAC, youth are at a higher risk of committing and falling victim to violence compared to other age groups, which has important and negative consequences for their life trajectories and for society as a whole. In her study using a panel of LAC countries, Chioda (2017) also finds that although the general level of employment was not

[5] Barbados, Antigua and Barbuda, St. Vincent and the Grenadines, St. Lucia, St. Kitts and Nevis, Commonwealth of Dominica, Grenada, Guyana and Trinidad and Tobago.

related to crime levels, youth unemployment was, meaning that higher levels of youth unemployment translate to higher crime levels.[6]

For the Caribbean region, in particular, youth development workers interviewed for the 2013 Social Impact youth assessment (Foss et al., 2013) reported that besides economic pressures, "deteriorating social support systems, lack of positive male role models, and the growing number of unattached males in an age group that tends to be less risk-averse combine to create opportunities for disorder and crime" (Foss et al., 2013, p. vii). One reason cited for increased gang membership in Guyana was the high number of parents who live elsewhere due to better economic opportunities, thus leaving their children with a single parent/caretaker or even with an older sibling. As household or family ties fail, youth seek to form other ties, including by engaging in promiscuous behaviors and joining gangs. Also, in Guyana, almost half of interview respondents characterized crime as incidents related to getting "fast cash" to fulfill basic needs. In terms of organized crime, one third of interview respondents said that youth are often used by adult perpetrators as front-line workers in drug trafficking and prostitution rings (Foss et al., 2013).

By interviewing youth in the study countries, Katz and Nuño (2017) were able to identify the risk factors and reasons for joining what they term troublesome youth groups (TYGs). Among the risk factors found to be significant, they list low self-control (all countries), low commitment to school (St. Lucia), negative peer commitment (St. Lucia and Guyana), and lack of parental supervision (Guyana). The most commonly cited reasons for joining TYGs were to make friends and because someone in the family was a group member. In Guyana, another reason was to participate in group activities. One stakeholder explained it this way: "Generally speaking . . . persons do not 'join' a gang, per se. It is simply a continuation of the community socialization process. Gangs are found in communities where many of their members reside; therefore, they are known to most persons in the communities and [have] socialized with them at some point. Being part of the group simply requires one to 'lime' with the gang in their customary location" (Katz and Nuño, pp. 12-3).

WHAT WORKS TO PREVENT YOUTH VIOLENCE

Most of the evidence concerning what works to prevent youth violence comes from the United States and other high-income countries. This section of the literature review begins by discussing findings on what works to prevent youth violence in general, based on studies conducted mostly outside the LAC region, and then focuses on the studies and reports that have been produced using LAC policies and programs. The evidence for Latin America is scarce, and even fewer studies have been identified for the Caribbean region. Finally, we look separately at the evidence concerning family-based therapy and counseling interventions, which is the area of work where the *Family Matters* program falls in.

According to a 2015 Inter-American Development Bank (IDB) study entitled *Closing Knowledge Gaps*, youth violence prevention strategies can be classified by their social ecology: "treatment-specific (i.e. cognitive behavioral therapy (CBT), counseling and social skills training), family-based (i.e. behavioral parent training and home visitation), and school- or community-based (i.e. mentoring and after-school programs, as well as social capital building outside of the family)" (Jaitman and Guerrero Compeán, 2015, pp. 16-7).

In a very thorough review of the literature on violence in Latin America and policies targeting children and youth, Chioda (2017) found that adolescence and young adulthood are critical ages for policy intervention "because of loosening of parental control, the increased pull of peers, the transition from

[6] Micro-level data for Mexico and Brazil also revealed that it was not just the employment status that matters for the relationship between labor markets and criminal offending, but also the quality of employment.

youth to adult roles, and the development of three different regions of the brain, which are responsible for the regulation of automatic or instinctive reactions, risk-taking behavior, self-control, and reflective reasoning" (p. 34). The most effective behavioral interventions identified by Chioda's review are CBT[7], life skills training and family-based interventions.

A recent and extensive literature review of over 264 studies, including only systematic reviews and individual studies that met the high standards of using randomized, experimental or quasi-experimental methods with an appropriate control group, found that the six types of interventions with the greatest evidence of effectiveness on preventing at-risk individuals and offenders from engaging in criminal and violent activities are: cognitive behavioral therapy, multidimensional therapy, drug courts and drug addiction treatment, focused deterrence, controls on the sale and abuse of alcohol, and hot spots policing (Abt et al., 2018). The first three types of interventions focus on the individuals that are already engaging in criminal and violent activities and seek to change their incentives and cognitive biases to achieve behavioral change. The focused deterrence approach focuses on the people that are high-risk and chronic offenders, and also seeks to alter the course of events by providing sanctions and rewards.[8] Controls on the sale and abuse of alcohol is the only kind of intervention not targeted at anyone in particular that has proven to be effective at reducing violence. There is clear evidence and consensus that hot spot policing can decrease crime rates in the targeted areas, but more recent research has questioned whether a displacement of crime to other areas is responsible for these decreases (Abt et al., 2018).

This literature review also identifies a category of promising interventions for which there is incomplete evidence of effectiveness: vocational training and employment, restorative justice, alternatives to incarceration, and conditional cash transfers. There is a third and much larger category of interventions for which there is inconclusive or contested evidence of effectiveness: crime prevention through environmental design, community-based prevention programs, school-based programs, and others. There are also some interventions where the evidence makes it clear that they do not work or in some cases can even be harmful, namely hospital-based prevention programs, bootcamps for youth offenders, "Scared Straight" programs, juvenile curfews, drug law enforcement, and gun buyback programs (Abt et al, 2018).

---

[7] Cognitive behavioral therapy (CBT) is a short-duration intervention that helps participants recognize and mitigate unhelpful thoughts and behaviors, or in a way, gets them to be "thinking about thinking" (Chioda, 2017). Participants learn to recognize situations where automatic, intuitive decision-making kicks in and results in negative outcomes, as a means of prevention. The CBT-based becoming a Man program for minority male youth in Chicago received considerable attention from the Obama administration and the press after an impact evaluation showed very positive results. Participation in the program raised an index of school engagement and performance by 0.14 and 0.19 standard deviations in the program and follow-up years, with a resulting 10-23 percent increase in the graduation rate relative to the control group. Program participants also experienced a 44 percent reduction in violent crime arrests and a 36 percent decline in arrests related to vandalism and weapons crimes during the program (Chioda, 2017). CBT has been successfully implemented in other cities in the United States.

[8] Focused deterrence is also known as the "Pulling Levers" strategy, and consists of: 1) selection of a crime problem, usually youth or gun homicide; 2) identify and analyze they key group of offenders and their behaviors; 3) assemble a multi-sector task force that includes the police, service providers and community representatives; 4) target these groups for special enforcement operations, aiming to alter their behavior by any legal means necessary; 5) supplement the law enforcement work with legitimate offers of assistance to these groups, and engage the "moral voice of the community:; 6) communicating clearly, often, and face-to-face with these groups to let them that they are under scrutiny, and their actions will have either positive or negative consequences, individually and for the whole group. Therefore, this strategy reaches out to offending groups directly, tracks and communicates with them clearly and often, and by "pulling every lever" with credible threats of enforcement and assistance, it seeks to alter their behavior. This intervention also has a clear focus, for instance on reducing the number of shootings, rather than seeking to stop all crime or convince members to leave their gangs. In Chicago, participants were 23 percent less likely to be involved in shootings and 32 percent less likely to become a gunshot victim in the year after treatment. Focused deterrence has been successfully implemented in other cities in the United States as well (Abt and Winship, 2016).

Another comprehensive study, a meta-review that examined 30 different strategies, also provides important insights and lessons on what works in reducing violence among youth (Abt and Winship, 2016). The meta-review identified mostly interventions that took place in the United States, and the authors complemented the work with fieldwork to countries in the Northern Triangle to better understand how these interventions could be applied there. The authors identify program elements of effectiveness, common among successful interventions: focusing on the people, places and behaviors that are most at risk for violence, being proactive in preventing violence, working together with partners and creating a legitimate loop between formal (police) and informal social control (communities), having good capacity in terms of resources and good implementation, and putting forward a clear theory of change (Abt and Winship, 2016). Of the 30 different strategies examined in their review, two stood out as particularly successful: focused deterrence and CBT.

According to a 2015 World Health Organization (WHO) report, the most promising interventions to prevent youth violence are implemented at the community level, such as community-police partnerships, reducing access to firearms, and drug control programs. Among school-based initiatives, the authors identify life and social skills development and bullying prevention programs as promising initiatives, but the effectiveness of other types of school-based initiatives is less evident (WHO, 2015).

School-based interventions can be successful, particularly when they are designed to alter behavior, thinking patterns, attitudes and beliefs. Those that discourage students from dropping out during their final years of secondary education yield benefits even decades later, in the form of reducing participation in violent crime and crime against property (Chioda, 2017). The YMCA Youth Development Programme (YDP), implemented in schools in Jamaica, is one such school-based intervention. A type of LifeSkills Training (LST) program, the YDP received the first evaluation for a program of this kind in a developing country context. YDP is a three-year intervention for middle and junior high school students, targeting violence as well as tobacco, alcohol and marijuana use. Youth receive daily supervision from 8:00am until 2:00pm, and the program provides instruction, socialization and training in social skills and self-management (Guerra et al., 2013). LST was evaluated multiple times and has shown positive results in reducing risk factors the short and long-term, acquiring the "effective" program label by the Blueprints Initiative. Although participants in the treatment group did not differ from their counterparts in the control groups in terms of "triggers" of aggression, as measured by a propensity scale, they showed better self-control by being less likely to act on those triggers (Chioda, 2017).

In terms of more evidence for the LAC region specifically, a very recent review identified only one systematic review and three other single studies focused specifically on interventions delivered in the LAC region (Campie et al., 2019). The systematic review identified nine studies on preventing youth violence in LAC. The interventions studied were comprised of school-based violence prevention (4), community-based violence prevention (3), parenting programs (1), and policing practices (1). This recent review mostly identified that there are large gaps in evidence for what works to prevent youth violence in LAC, and that studies often have significant limitations. For example, 1) no longitudinal study was identified, which would have been able to detect changes over an extended period of time, 2) community-level outcomes are usually measured instead of individual ones, which makes it more difficult to determine if any changes are attributable to those directly involved in the program, 3) sometimes the same individuals implement and evaluate the intervention, which raises concerns, 4) many RCTs had small sample sizes without a clear indication of whether the study had enough power to detect small effects, among other limitations (Campie et al., 2019).

Although most of the studies included in the systematic review found positive or promising results, particularly the community-based programs, it concluded that the evidence for Latin America is scarce and relies on non-rigorously designed studies (Atienzo et al., 2017). In addition, it highlighted the challenges of transferring an intervention from abroad into the Latin American context, where there are "high levels of inequality and poverty, a lack of quality education, a culture of masculinity that promotes the involvement in conflict, urban growth and a drug-trafficking context" (Atienzo et al., 2017, p. 16).

While many interventions have been studied in isolation, a large evaluation published in 2014 measured the effect of various USAID's efforts on combatting crime and violence in Central America. This is a multi-year, multi-method and multi-country evaluation. The list of activities and programs evaluated included municipal-level committees, crime observatories and data collection, environmental design projects (such as street lighting and cleaned up public spaces), programs for at-risk youth, and community policing. Both quantitative and qualitative data pointed to a significant improvement in crime reduction in treatment communities. Large, statistically significant treatment group improvements include smaller numbers of: reported murder cases (51 percent), extortion (51 percent), people avoiding walking through dangerous areas (35 percent), sales of illegal drugs (25 percent), robberies (19 percent), and report of young people in gangs (14 percent). Notably, criminal activity is measured via survey, not via reports to police (Berk-Seligson et al., 2014).

FAMILY-BASED THERAPY AND COUNSELING

Developmental research findings suggest that the family environment includes key risk factors for involvement in crime and violence, namely household structure (single-parent households, siblings as caregivers, etc.), parents' age at first birth, and the stability of the home environment (Chioda, 2017). Family history and criminal background, particularly of the parents, are among the strongest predictors of involvement with criminality, even stronger than income or employment status (Chioda, 2017). Gang membership has also been shown to be a significant risk factor for offending and involvement in delinquency (see discussion in Thornberry et al, pp. 955-6). Improving parent-adolescent communication can lead to better mental health, better social functioning and more positive behavioral outcomes (UNICEF, 2017).

Therapeutic approaches "include cognitive, behavioral, psychosocial or social interventions delivered to individuals who have been involved in aggressive or violent behavior, or who are at risk of committing such behaviors" (WHO, 2015, p.39). These interventions may be delivered in an individual or group setting by social workers or trained therapists and may or may not include the family.

The most common types of therapeutic interventions that seek to help families function better include functional family therapy (FFT), multisystemic therapy (MST), and brief strategic family therapy. Through therapy, these interventions seek to replace dysfunctional family behaviors with healthier ones to improve antisocial behavior in youth. Participating families will have one or more youth who are struggling with serious antisocial behavior or are already in trouble with the law and will work with a trained therapist over the course of three months or longer. Each of these three types of family-based therapies has been evaluated multiple times and was found to reduce incidences of anti-social behavior, arrests and recidivism (Chioda, 2017). The WHO report highlights that a meta-analysis of FFT found small to medium effects on the recidivism of juvenile criminal offenders. Therapeutic approaches that include cognitive behavioral therapy seem to work best in reducing youth engagement with crime and violence (WHO, 2015).

Systematic reviews of MST and FFT suggest that these interventions are effective at reducing engagement in criminal activity (see Abt et al., 2018 for a complete list of systematic reviews and studies included in their report). Model adherence appears to be very important for these highly structured interventions, with their effectiveness being dependent on the therapists adhering to the model (Sexton and Turner, 2010). One study that reviewed eight RCTs in the United States found that MST, a type of intensive family and community-based treatment, had no statistical effect on arrests and convictions (Littell, Popa, and Forsythe, 2005). However, this review includes many unpublished and less rigorous studies, so the results should be interpreted with more caution (Abt et al., 2018).

A more recent RCT of an FFT program in Philadelphia found that the effectiveness of such a program on recidivism varied by levels of risk joining a gang. Among youth that were at a high risk for joining a gang, the study found significantly lower recidivism rates, compared to the control group that was receiving the "treatment as usual." For youth that were at a lower risk for gang membership, the study did not identify consistent differences between the treatment and control groups (Thornberry et al, 2018).

The *Family Matters* program is based on the family counseling pillar of the Gang Reduction and Youth Development (GRYD) program developed in Los Angeles. GRYD includes six interrelated approaches: 1) primary prevention (for instance, a Gun Buy-Back program and community forums where GRYD staff present information on gang risk factors), 2) secondary prevention (targeting youth ages 10-15 who are assessed as high risk), 3) family case management (youth ages 14-25 who are engaged in gang activity), 4) incident response (immediate response to gang-related violent incidents, to control rumors and mitigate tensions that might lead to further issues such as retaliatory violence), 5) community engagement (engage the community and police in a community policing capacity, and incorporating community engagement into all of GRYD's work[9]), and 6) suppression (the GRYD office continually communicates with law enforcement). Therefore, this program includes family-based programming but also five other pillars seen as essential to violence prevention (Cahill et al, 2015). Another way to summarize the GRYD areas of work are community engagement, gang prevention (where the family counseling services fall in), gang intervention, and violence interruption (Kraus et al, 2017).

A 2015 evaluation of the GRYD family counseling component found that youth clients and parents improved family functioning and problem-solving over the course of the study. The process evaluation clearly showed very important shifts in behavior and perception. However, when assessing changes in community outcomes that could be attributed to the overall GRYD intervention, the study had mixed findings. While there were large reductions in the incidence of crime and violence and a shrinking of crime hot spots, these general trends were not limited to the treatment communities. The evaluation found no statistically significant differences between treatment and control communities that would suggest programmatic impact (Cahill et al, 2015). Hence, although the evaluation generally showed encouraging findings in terms of reported improvements by the participant youth and the community, it is not altogether clear that crime levels changed as a consequence of the program.

A more recent evaluation of the GRYD family counseling component was published in 2017 (Kraus et al, 2017). Treatment youth were surveyed with the YSET right before the counseling started and after six months of intervention. Youth in the program experienced large risk reductions across all nine YSET scales, and 57.5 percent of them were no longer at risk and no longer eligible for the program after those six

[9] CFYR is also doing some of this engagement with Community Enhancement Committees in St. Lucia and Guyana, and up until August 2018, also in St. Kitts and Nevis, but this was cancelled following program budget cuts.

months (as measured by scoring at-risk on four or more YSET scales). While these are large changes, we cannot say with certainty that a comparable control group would not have experienced similar changes in YSET scores. This evaluation was not an RCT, and the control group used in this study appears to be different from the treatment group in significant ways. The comparison group for this evaluation was composed of high-risk youth on juvenile probation in Los Angeles, and most were referred by Deputy Probation Officers. Youth took the YSET survey, and those that scored at-risk were used as a comparison group for GRYD Prevention Services clients. While the baseline scores were very similar for the two groups, youth in the comparison group differed in some significant ways, most notably in age (older), gender (more males) and ethnic composition (more Latinos and fewer African Americans). The evaluation report does not include tests of sample balance to check that the two groups are indeed comparable enough. In addition, matching was not conducted on characteristics to create a more comparable control group. Therefore, we are not confident that the differences found between treatment and control in this evaluation are indeed caused by the intervention and not by other factors that were not controlled for.

The *Proponte Más* program in Honduras was also based on the family systems secondary prevention model implemented in Los Angeles and designed by Creative Associates. By strengthening family relationships and youth support and supervision, the program sought to build resistance to gang recruitment. Individual and group activities were carried out through seven phases to promote change in youth high-risk behaviors inside and outside the family environment. Although the *Proponte Más* program is similar to *Family Matters* in terms of the structure of the family counseling provided to participants, there are important differences. For instance, *Proponte Más* worked with referrals instead of blanketing the communities and visiting schools to implement YSETs, as was done for *Family Matters*. In addition, with *Proponte Más,* family counselors applied the YSETs instead of independent enumerators.[10]

An impact evaluation of the second cohort of the *Proponte Más* program (following a pilot) was implemented in 2018 (Katz, Cheon and Decker, 2019). The YSET was used at baseline and six months later to test for changes in risk factors. There were improvements in risk levels in both the treatment and control groups, and in most of the nine scales, the treatment group's improvement is not statistically significantly different from that of the control group. The exceptions are the scales A (Antisocial tendencies), B (Weak Parental Supervision), and DE (Impulsive risk taking). For the outcome variable "eligible" (i.e., scoring at-risk in four or more risk factors) the difference between treatment and control is not statistically significant. The evaluators found more positives changes in the Family cohesion index questions that are asked of caregivers, although they acknowledge that the effect sizes are small.

[10] These differences between *Proponte Más* and *Family Matters* occurred because a rigorous and well-powered impact evaluation of *Family Matters* required a larger sample size from the outset (with blanketing of communities instead of working with referrals), and that enumerators instead of family counselors applied the YSET to avoid social desirability bias on the part of youth respondents.

BACKGROUND ON THE CFYR FAMILY MATTERS PROGRAM

USAID/CFYR's *Family Matters* intervention (*Family Matters*) is a secondary violence prevention method that targets youth between 10 and 17 years of age who are at a secondary or tertiary level of risk of engaging in delinquent behaviors leading to crime and violence. Eligible youth are identified through the Youth Services Eligibility Tool (YSET) assessment, used to determine their level of risk. The intervention is grounded in research that shows that positive behavioral changes in youth are more likely to last if embedded in and reinforced by the family and the larger community.

Throughout the implementation cycle, family counselors trained by USAID/CFYR in the *Family Matters* methodology hold regular face-to-face counseling sessions with participating youth and their families. During these meetings, typically held in the family's home, counselors strengthen family cohesion (bonds that hold family members together) and connect the family to the wider community. Family counselors also work with youth on an individual basis to help them adopt positive and safe behavior. By the end of the cycle, family members are expected to interact more frequently, communicate more effectively, make joint decisions more regularly, demonstrate a stronger sense of family pride, and exhibit a stronger reliance on community networks for support. These improvements help families better protect and guide youth.

The secondary prevention approach is grounded on family system theory and practice, including Structural, Multigenerational, Communications and Strategic approaches. The theory of change postulates that all the risk factors and associated behaviors are reinforced by the relations within the family. If family member behaviors and practices associated with risk factors can be identified and modified, family dynamics and relations should improve and risk factors and associated negative behaviors should be reduced (CFYR Family Counseling Training Manual, 2017). Other factors that have been identified as being important for improving anti-social behavior include increasing youth's resilience factors, such as a more stable home environment, the presence of positive male role models, higher self-esteem, and peaceful conflict resolution skills, while decreasing overall risk factors, such as low literacy levels, unemployment, domestic violence, and child abuse.

Family Matters engages beneficiary youth and their families in an approximately seven-month cycle of structured family counseling, specifically adapted for the Caribbean context. Counselors work with families for a total of two cycles, which add up to 12-14 months of intervention. Figure 2 lists all of the phases in each cycle. The difference between the first and second cycle is that in the second cycle, the family is expected to take charge based on what they learned during counseling and focus on other youth and family behaviors that may still need to improve. Phase 1 denotes the referral and collaboration process, which in the case of this program entailed blanketing the treatment communities to try to survey all the youth within the age range. The program was designed to work through referrals, but because of minimum sample size requirements for the impact evaluation, this phase included blanketing the communities.

Figure 2: *Family Matters* phases

Source: CFYR brochure.

Before the intervention started, youth in target communities were surveyed using the YSET. CFYR and its partners at USC identified youth risk across nine YSET modules and determined eligibility for the intervention. Youth were eligible if they scored at-risk on any four (or more) of the nine main YSET modules (See Annex I for more details about how cut-points for each module were determined). Out of the nine scales used for risk assessment, seven are attitudinal and two are behavioral, as listed below (Kraus et al., 2017).

Attitudinal scales:

- A: Anti-social tendencies
- B: Weak parental supervision
- C: Critical life events
- DE: Impulsive risk taking
- F: Guilt neutralization
- G: Negative peer influence
- H: Peer delinquency

Behavioral scales:

- IJ: Self-reported delinquency
- T: Family gang influence

The program is based on the PIFSM, but it was adapted to the Caribbean and further adapted to allow for an impact evaluation. For example, in the PIFSM, counselors conducted the initial YSET assessments to help build a trusting relationship with the youth and families. However, from an evaluation perspective, having the same counselors applying the YSET and providing counseling would undermine the validity of the YSET as an evaluation tool, particularly at endline. As such, the YSET was applied by researchers trained with CFYR's

engagement instead of counselors. It is possible that removing this trust building function undermined CFYR's ability to attract program participants and build initial trust.

The implementer selected communities by first conducting a quantitative analysis of crime and offending data and demographic data which provided an initial ranking of communities. Qualitative research and site visits were conducted to further refine the list. Few communities changed their position in the rank order following the use of qualitative tools to assess community need and appropriateness. Table 1 lists the 15 communities selected for all CFYR implementation activities, which were among communities with higher levels of crime and violence. Five communities from each country were selected, and all 15 are included in the evaluation. CFYR then identified schools which educate youth from each of the selected communities in order to source youth for the program. Within each school, CFYR conducted classroom visits to determine which youth were from treatment communities. After attaining parental consent for these students, enumerators interviewed each student using the YSET tool. CFYR identified 16 schools in Guyana, 17 schools in St. Kitts and Nevis and 30 schools in St. Lucia.

Table 1: List of countries and communities selected for the intervention

Country	Communities
Guyana	Sophia (Fields C, D and E), Lodge, East Ruimveldt, East La Penitence, and Corriverton
St. Lucia	Castries (Bois Patat, Morne du Don, Grass Street, Leslie Land and Marchand), Anse La Raye, Soufrière, Vieux-Fort and Dennery
St. Kitts and Nevis	In St. Kitts: Basseterre (The Village, McKnight and Newtown), Cayon, and Sandy Point. In Nevis: St. George and St. John

EVALUATION PURPOSE, QUESTIONS, AND METHODOLOGY

EVALUATION PURPOSE

This study employs an RCT to measure the impact of the CFYR *Family Matters* program on participating youths' overall risk and risk across nine risk factors, as detected by the YSET tool. The evaluation team also designed a caregiver survey to explore the impact of the intervention on families and to collect additional information to better understand which kinds of families may benefit more from the program, and why.

EVALUATION QUESTIONS

The evaluation seeks to answer two core questions:

1) Do secondary prevention efforts impact the participating youth's overall risk, crime and violence-related behavior, and their individual risk and resilience factors? Why or why not?

2) Do secondary prevention efforts improve the family environment? Why or why not?

To answer the first question, the main data source is the YSET collected at baseline, midline and endline. To answer the second, the main data source is the caregiver survey, collected at baseline and endline. In addition, we have used a community survey that CFYR collected at two points in time in each of the participating communities, to check for any community-level changes that could be affecting the evaluation results. The YSET, caregiver and community instruments are included in the supplementary Annex IX.

METHODOLOGY

This evaluation examines the impact of the *Family Matters* program on youth risk using a multi-site RCT to isolate causal effect. The study randomizes youth considered to be at a high-risk level based on the YSET assessment tool. Enumerators identified households with youth ages 10 to 17 in the 15 communities across St. Lucia, St. Kitts and Nevis and Guyana, and then implemented the YSET tool after obtaining parental and youth consent. Using risk thresholds for nine of the YSET modules, CFYR and their partners at USC determined if youth were at high risk and thus eligible for the program (which was based on a score of four or more modules where the youth scored "at high risk.") The evaluators then randomized eligible youth into a treatment group, who were offered the program for the 2019 year, and a control group, who were not offered the program in 2019 but would be offered the program in 2020 at the conclusion of the evaluation.

The YSET tool was used again after six months (midline) and after 12 months (endline) to survey youth who were randomized to the treatment and control groups. By comparing the youths' answers from baseline to endline, we can find out if the intervention had a positive impact on their behavior and attitudes. Because the youth's behavior and attitude may change over this period for other reasons that are not related to the intervention, we have a control group that is as similar as possible to the treatment group, with the only difference being that the control group did not participate in the program. With a comparable control group and a sufficient sample size, we can be confident that any differences we see between treatment and control over time are caused by the intervention.

In addition, we conducted a survey with the primary caregiver of each of the eligible youth randomized into the control and treatment groups. This allows the evaluation to test the effects of the intervention on the family and obtain the perspective of primary caregivers.

While a comparison of randomly assigned youth will allow us to estimate the counterfactual of not having participated in the program, there are several challenges and limitations that should be mentioned. First, the effect is measurable as long as the study has enough statistical power,[11] a function of the number of youth in the program, their geographical distribution, and variation in key indicators. While the study's power is not sufficient to examine heterogeneous effects across all subgroups of interest (e.g., the impact of the program in an individual community), it is large enough to test overall program impact and some heterogeneous effects. Second, the impact evaluation requires that the initial assignment of youth into treatment and control be respected, and because of good implementer cooperation, assignment was followed. Third, the evaluation requires that the YSET and caregivers survey accurately and reliably measure intended outcomes and changes over time. This is not a small challenge, as youth and caregivers might not respond to surveys accurately or reliably. We discuss this issue in greater detail as we interpret the results. Fourth, non-participation in the program and attrition from the program and the study could undermine the balance between treatment and control groups, and this did occur. We employed two measures to address this concern. First, we use a difference-in-difference approach to data analysis, which does not require baseline balance, and second, we draw a smaller sample from the larger control group to create a matched comparison group and achieve greater balance between the treatment and control groups. Because the study successfully employed randomization of youth into treatment and control youth, we can reasonably say that the parallel trends assumption is met. In the following sections, we discuss the difference-in-difference estimation strategy and then the matching.

ESTIMATION STRATEGY

Before starting with the midline and endline analysis, we wrote a pre-analysis plan of our estimation strategy, shared with USAID and Creative/CFYR, and registered it with Evidence in Governance and Politics (EGAP), to be gated until 3/1/2021.[12]

Using the difference-in-differences framework in analyzing the YSET and the caregivers survey, the evaluation essentially measures the extent to which the mean is different between the treated and control groups of youth and caregivers, between baseline and endline. Although this seems like a simple test, we should recall that a hypothesis test requires having information on both the expected value of the quantity of interest as well as its dispersion, usually measured through the standard deviation. The simplest test of means can be done through an ordinary least squares (OLS) regression model, comparing the average risk as measured by some indicator from the YSET, after finishing the program, for youth that have been given counseling through CFYR, compared to equivalent youth that were not in the program, from a starting condition of similar risk they exhibited at the beginning of the program. That OLS estimation is not completely correct, however, because it does not take into account that the observed youth are repeated, in the sense that they are observed at two moments in time (through baseline and endline surveys). Hence

[11] We check for the MDEs in the YSET and caregiver analysis sections. Our conclusion is that the evaluation has enough power to detect even very small effect sizes, and that failure to detect a statistically significant effect means that there is no meaningful relationship between the program and the outcome in question.

[12] The pre-analysis plan registration was later migrated by EGAP to the OSF registry, and can be found via this link: https://osf.io/k85xn/

a more accurate measurement of the true variance is one that acknowledges that a youth behavior measured in the past may condition their behavior at the end of the program. A generalized least squares (GLS) random effects estimator produces a weighted average of the effects of the program, recognizing that these are made up of effects between youth in the cross-section, as well as within effects of an individual youth changing through time. This produces a slightly different mean that can be calculated through estimations where the errors are clustered by youth, recognizing that individual observations of the same person before and after the program are not strictly independent.

We also fine tune estimates by measuring the effects in what is usually referred to as a "fixed effects" model, in which the regression estimator shows the results as the mean within youth. However, this model is not always appropriate, and we compare the random effects and fixed effects models through a Hausman test to find out whether there is a systematic difference between two estimators. If the fixed effects estimator fails this test, in the sense that there is no difference in the coefficients, the random effects estimator is preferable, since it has the desirable feature of being efficient and consistent.

A final consideration is that even though the GLS fixed effects estimator may not be the preferred one, there are substantive reasons to include fixed effects in the estimation generated by the design of the program and the evaluation. Since the youth are clustered into specific communities from each of the different countries, there might be differences in the change of risky behaviors or attitudes across communities and countries that are not attributable to the program. Such effects should not be included as a program effect emerging from the intervention.

The evaluation employs a number of estimation strategies, including 1) simple difference-in-differences, 2) with random effects and clustered standard errors at the individual level, 3) with random effects and clustered standard errors at the individual level, and community dummies, and 4) with fixed effects and clustered standard errors at the individual level. We primarily rely on the third approach, difference-in-differences estimators calculated through a GLS random effects model, with variance corrected errors clustered at the level of individual youth, including fixed effects for the 15 communities in the study; however, we use all four models to ensure that we are not presenting model-dependent results. The errors calculated in this way are slightly larger than those that would emerge from a simple test of means or an OLS estimation; however, we are confident that such an approach provides efficient, consistent and unbiased estimators of the true size of the effect of the program. Nonetheless, for ease of interpretation, in the body of the report we tend to present simple comparisons that the reader can follow without sophisticated methodological knowledge and then note any differences with more accurate models.

The primary specification to identify the impact of the program on youth risk is:

$$y_{iv} = \beta_0 + \beta_1 \mathrm{TREAT}_{iv} + \beta_2 \mathrm{TREAT}_{iv} * \mathrm{F} + \beta_3 \mathrm{F} + \beta_4 X_{iv} + \mathrm{m}_v + \epsilon$$

Where y_{iv} is the outcome of interest; TREAT is the treatment condition (0=control, 1=treatment); F is the follow-up period (0=baseline, 1=endline); X_{iv} are individual level characteristics, such as age, gender or school status, that may not have achieved balance; and m_v is the fixed effect of each community. β_2 measures the impact of the intervention on the outcome of interest y_i.

Before running the difference-in-difference analysis, we check the balance between the treatment and control groups on various socio-demographic and risk variables, to make sure that we are comparing youth who have similar observable characteristics.

We also conduct the analysis on two different populations. We primarily focus on what is referred to as the treatment-on-the-treated (TOT), and also known as local average treatment effects (LATE). This analysis compares treatment youth that participated in the program to those that were randomized into the control group. However, we also conduct an intention-to-treat (ITT) analysis. This ITT approach compares all youth randomized to treatment, regardless if they participated in the program or not, and youth randomized to control. The ITT analysis measures the intended impact of the intervention on the *population* of interest rather than on those that participated in the program.

MATCHING

As noted above, we also address the potential effects of non-participation and attrition through a matching approach. For the exact matching process[13], we use baseline data only and randomly select a reduced sample size for the control group, which is 33 percent greater than the treatment group in each site. The reason for selecting 33 percent is that, at that point, we start to run out of control youth in some of the sites (Cayon, Sandy Point, and Anse la Raye). We could select a larger control group sample, but then communities will be less and less balanced, so there is a trade-off between community balance and power.

After the matching algorithm draws a random sample, it also runs balance tests on the following baseline variables: age, gender, school status, country, community, YSET score, score on each of the nine main YSET modules, and module risk variables, which are binary, with one equaling at-risk for that module (these outcome variables will be explained in greater detail later in Table 4). Therefore, we check for balance between treatment and control in several socio-demographic characteristics and in all of the main risk variables. If the sample does not balance in all of these, then the algorithm goes back to the beginning and draws another sample of control. This matching process continues until it finds a sample that balances. In order to be considered balanced, two-tailed t-tests must have a p-value of 0.2 or higher for all of the aforementioned variables when comparing the treatment and control sample means. This technique ensures that we are comparing youth that are as similar as possible, given the data that we have available; however, it does reduce the sample size.

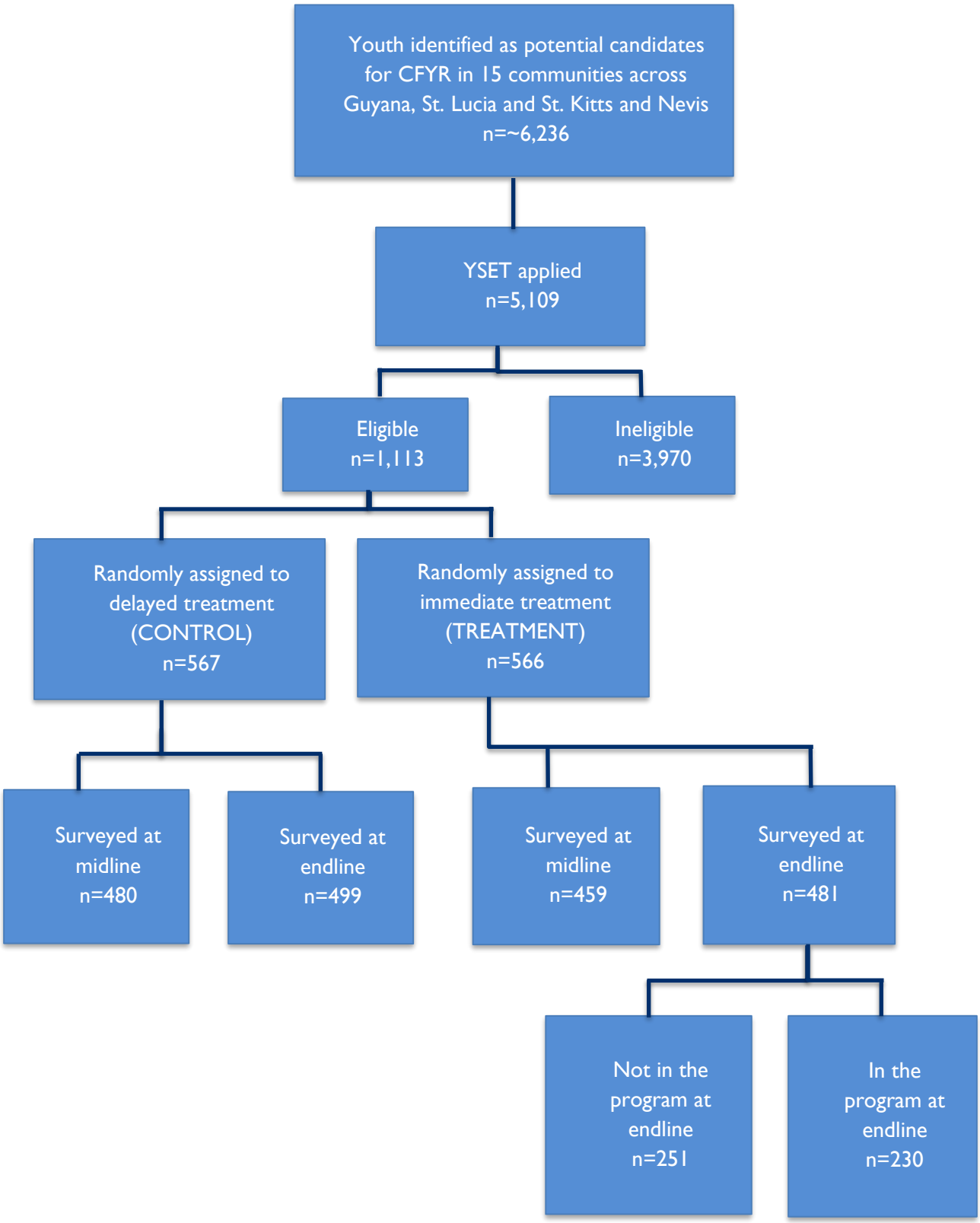
[13] We considered two matching strategies: coarsened exact matching (CEM) and exact matching. Exact matching is the preferred method because CEM reduced the sample size of the treatment group and yielded imbalances on the community variable and some outcome variables.

SAMPLE DESCRIPTION

YOUTH SAMPLE DESCRIPTION

The flow chart in Figure 3 shows the number of youth considered for the program over time. Over 6,000 youth between the ages of 10 and 17 were identified by CFYR, SI, and NORC in the 15 communities studied, and enumerators were able to successfully interview 5,109 of them using the YSET tool. Of those surveyed, approximately 22 percent (1,113) scored above the risk threshold on four or more risk modules in the YSET, making them eligible for the program. Of the eligible youth, 566 were randomized to treatment and 567 were randomized to control. Enumerators attempted to survey all treatment and control youth again at midline and endline, but some refused or could not be located. While somewhat higher in the treatment group than in the control, survey attrition was limited to a reasonable 14 percent of the total randomized sample, meaning that 14 percent of the original sample was not re-interviewed. The entire group that was surveyed at endline (481 in the treatment group and 499 in the control group) form the ITT sample. The youth that completed the program and were surveyed at endline compose the treatment arm of the TOT sample, while the control group is the same as for the ITT sample. Of the treatment youth surveyed at endline, 230 (48 percent) had completed the intervention by finishing the second cycle with their family counselor. The remaining youth had either dropped out (20 percent), declined the intervention (16 percent), or not been invited to participate (16 percent).

Figure 3: Number of youth considered for the program and surveyed over time



For most of the analysis in this report, we only include youth for whom we have baseline and endline data. Because survey participation was somewhat lower at midline, the sample of youth included in the analysis is somewhat smaller at midline than at endline. Table 2 displays the number of youth by data collection period, country and other categories that comprise the panel sample. The ITT sample includes all of those randomized to treatment who were re-interviewed at endline (481 youth). The TOT sample only includes those youth randomized to treatment who completed the intervention and were re-interviewed at endline; and therefore, it has fewer youth (230). Both samples have the same number of control youth (499). St. Lucia has the largest sample, with 98 treatment youth, Guyana has 77 treatment youth, and St. Kitts and Nevis has the smallest sample, with 55 treatment youth.

"False positives" are youth who were erroneously classified as eligible at baseline due to scoring errors, and "false negatives" are youth who were erroneously classified as ineligible at baseline due to scoring errors. The scoring errors for most of these observations were only identified after randomization had been completed, and as a result some of the "false positives" youth were approached for the intervention. In our analysis for this report, we used samples including and excluding the false positives to check for any significant differences in results and determined that inclusion of false positives does not impact the results. In some cases, the results are more statistically significant when we include the false positives, likely a result of a larger sample. All of the analysis presented here, therefore, includes the false positives.

Table 2: Number of YSETs completed by category and data collection period that comprise our panel sample

Number of YSETs by Category	Baseline	Midline	Endline
Randomized to control	499	451	499
Randomized to treatment (ITT)	481	428	481
Completed treatment (TOT)	230	224	230
Guyana treatment (completed)	77	76	77
Guyana control	193	163	193
St. Lucia treatment (completed)	98	98	98
St. Lucia control	219	208	219
St. Kitts and Nevis treatment (completed)	55	50	55
St. Kitts and Nevis control	87	80	87
False positives	45	41	45
False negatives	13	3	13

For the qualitative research trips conducted at midline, we hosted focus group discussions with youth and caregivers from several communities in each country. Our field coordinators invited youth and caregivers

from a randomly ordered list, attempting to get 10-12 confirmations. Participant attendance among those who confirmed was low, around 50 percent. Other qualitative work conducted includes observation of family counselor trainings, before the program began, which allowed us to confirm that counselors were well trained.

CAREGIVER SAMPLE DESCRIPTION

The caregiver sample is composed of mothers, fathers, guardians and other relatives or non-relatives responsible for the youth. Unlike youth, caregivers were only surveyed at baseline and endline. Table 3 describes the sample of caregivers surveyed, disaggregated by treatment status, country and false positive/false negative status. As with the youth sample, the caregiver sample can be categorized in two ways. First, the ITT sample is composed of caregivers whose youth were randomized to treatment, regardless of whether they were offered or took up program. Second, the TOT sample is composed of caregivers whose youth were randomized to treatment, were offered the treatment, and completed treatment at endline. The TOT control group is the same as the ITT control group.

The total sample at baseline was 1,002, with 501 caregivers randomized to control and 501 randomized to treatment. These caregivers form the ITT sample. Since not all youth randomized to treatment took up the treatment, there were only 236 treatment caregivers in the TOT sample at baseline. At endline, the total sample was 1,024, with 529 control caregivers and 495 treatment caregivers in the ITT sample, and 240 treatment caregivers in the TOT sample. The TOT treatment sample included 83 caregivers from Guyana, 103 from St. Lucia and 54 from St. Kitts and Nevis at endline. The caregiver sample also included 47 false positives and 12 false negatives.

Table 3: Number of caregiver surveys completed by category and period (baseline and endline)

Number of Caregiver Surveys by Category	Baseline	Endline
Randomized to control	501	529
Randomized to treatment (ITT)	501	495
Completed treatment (TOT)	236	240
Guyana treatment	84	83
Guyana control	204	207
St. Lucia treatment	102	103
St. Lucia control	220	237
St. Kitts and Nevis treatment	50	54
St. Kitts and Nevis control	77	85
False positive	47	47
False negative	12	12

The caregiver and youth samples differ for two reasons: 1) Caregivers may have more than one youth participating in the *Family Matters* program, so there are cases where one caregiver answers survey questions about multiple youth; and 2) Survey responsiveness for some caregivers and youth may not match, so there are caregivers who respond to the survey whose youth did not participate in the survey, and vice versa.

FINDINGS BASED ON THE YSET ASSESSMENT

SAMPLE BALANCE

At baseline, youth were stratified by community and randomly assigned to treatment and control groups. Additional measures were taken to ensure that the treatment and control groups were balanced across several key variables, including outcome variables. Because of survey and program attrition, however, the final samples are smaller than the baseline sample, and there is a risk that balance on key indicators could have been reduced. The ITT sample was affected by survey attrition and the TOT sample was impacted by both survey and program attrition. As such, we run balance checks to determine if control and treatment groups in each of the samples have observable differences across dimensions of heterogeneity that could impact differences in outcomes.

The complete tables with balance tests and their interpretation can be found in ANNEX III: BALANCE TESTS. We set a conservative p-value of 0.2 and find that the ITT sample looks balanced, not requiring matching of treatment and control to create more similar samples. In the case of the TOT sample, however, there is less balance on several categories, including the demographic variable for school status (i.e., youth is in or out of school). To account for these baseline differences, in addition to running the analysis on the unbalanced, unchanged sample, we also run it on a matched sample that we generate using an exact matching technique. These two analyses do not differ substantially, and while we include the matched analysis in

ANNEX V: ADDITIONAL YSET SURVEY RESULTS, we are confident that imbalances do not impact our conclusions.

OUTCOME CONSTRUCTION

Table 4 indicates the four different ways to measure changes in YSET outcomes: 1) Using module scores to measure changes in mean scores; 2) Using module scores to measure changes in the share of youth who are scoring at risk for each module; 3) Using all of the scores on the nine main YSET modules to measure changes in the average number of modules for which the youth scored at risk; and 4) Using all of the scores on the nine main YSET modules to measure whether the share of youth scoring "eligible" (four or more risk factors) has changed. We explore changes across all four types of YSET outcomes.

Table 4: Four different ways to measure changes in YSET outcomes

Concept	Operationalization	Indicators
1. Change in average scores for each of the modules	0 – 20 (maximum varies by module)	A: Anti-social tendencies (mean) B: Weak parental monitoring (mean) C: Critical life events (count) DE: Impulsive risk taking (mean) EG: Risky behaviors (count) F: Guilt neutralization (mean) FSH: Horizontal family (mean) FSV: Vertical family (mean) G: Negative peer influence (mean) H: Peer delinquency (mean) IJ: Self-reported delinquency (count) T: Family antisocial influence (sum)
2. Change in the percentage of at-risk youth for a given module	0-100: Percentage of at-risk youth	B: Weak parental monitoring C: Critical life events DE: Impulsive risk taking EG: Risky behaviors F: Guilt neutralization G: Negative peer influence H: Peer delinquency IJ: Self-reported delinquency T: Family antisocial influence
3. Change in the average number of risk factors (number of modules youth are considered at risk for)	0-9	Mean number of risk factors
4. Change in the percentage of at-risk youth (scoring at risk for 4 or more YSET modules)	0-100: Percentage of at-risk youth	Four or more risk factors (also referred to as "eligible" because that was the threshold for program eligibility)

The YSET survey is designed so that in each module and for the overall YSET, a higher score means a higher risk level. In ANNEX IV: OUTCOME VARIABLE CONSTRUCTIONwe explain how the module risk variables, which fall under concept number two in Table 4 above, were created. The baseline and endline instruments, with the score assigned to each question, can be found in the supplementary annex, Annex IX.

ANALYSIS RESULTS

Figure 4 displays the share of youth scoring eligible (scoring at risk for four or more YSET modules) by data collection period for the ITT and TOT samples, using a simple estimation strategy without any matching done to the samples. The error margins show the upper and lower score bounds at the 90

percent confidence level. As stated earlier, the control group is composed of the same youth in both the ITT and TOT samples.

At baseline, the share of youth scoring eligible for the treatment group is nearly identical to that of the control group.[9] From baseline to endline, the TOT and ITT samples trend in the same direction, with similar change sizes in both control and treatment groups. In the TOT sample specifically, the treatment group goes down from 94 percent at baseline, to 54 percent at midline, to 43 percent at endline, and we observe a similar trend for control youth. While the most salient aspect of the figure is the dramatic change in both treatment and control, we do find a slightly larger change among the treatment group in the TOT sample, and this "difference-in-difference" is statistically significant at the 90 percent confidence level. The full table with coefficients and levels of statistical significance for all of the main outcomes is presented later in this chapter, in

Table 6 and

Table 7.

Figure 4: Percent of youth scoring eligible (four or more risk factors), by treatment status, data collection period and sample (with standard error bars)

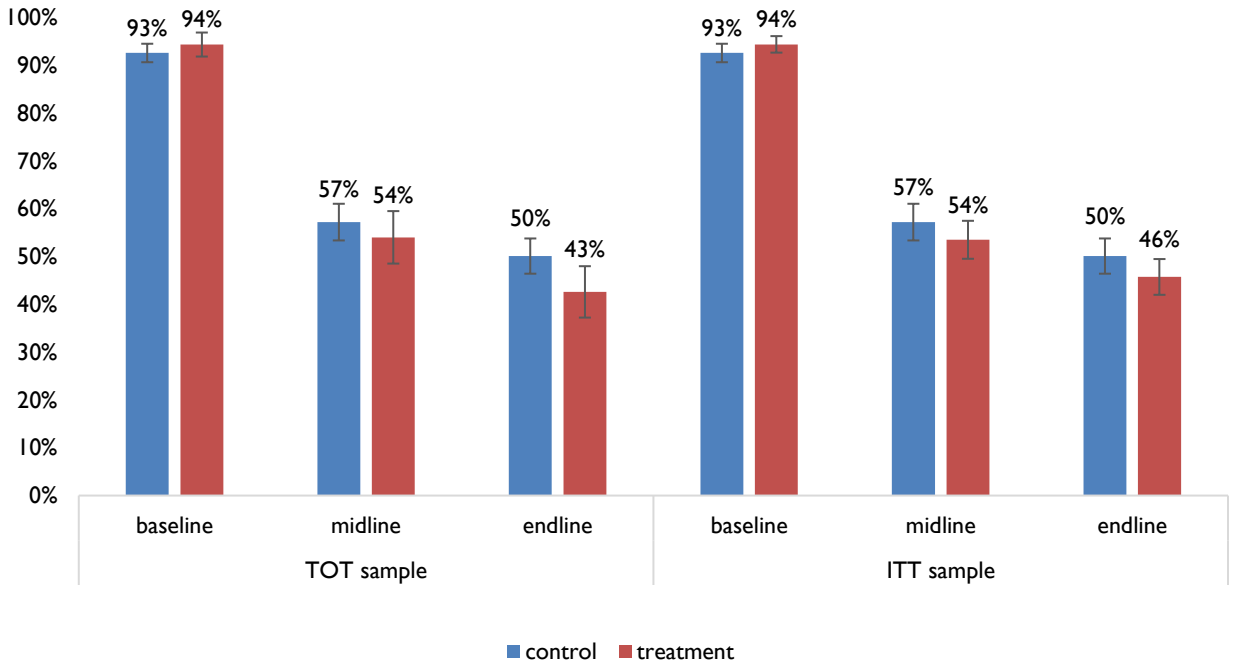


Figure 5 also shows the share of eligible youth by data collection period and treatment status, but disaggregated by country, and only for the TOT sample. While the treatment and control groups present similar shares at midline and endline in St. Lucia and St. Kitts and Nevis, in Guyana, the difference between treatment and control is statistically significant at endline, suggesting that the intervention might have had an impact in this country. While this simple comparison pools Guyana youth and does not include a control variable for community-level effects, as required by best evaluation practice, we find similar results in more sophisticated analyses below.

Figure 5: Percent of youth scoring eligible (4 or more risk factors), by treatment status, data collection period and country for the TOT sample (with standard error bars)

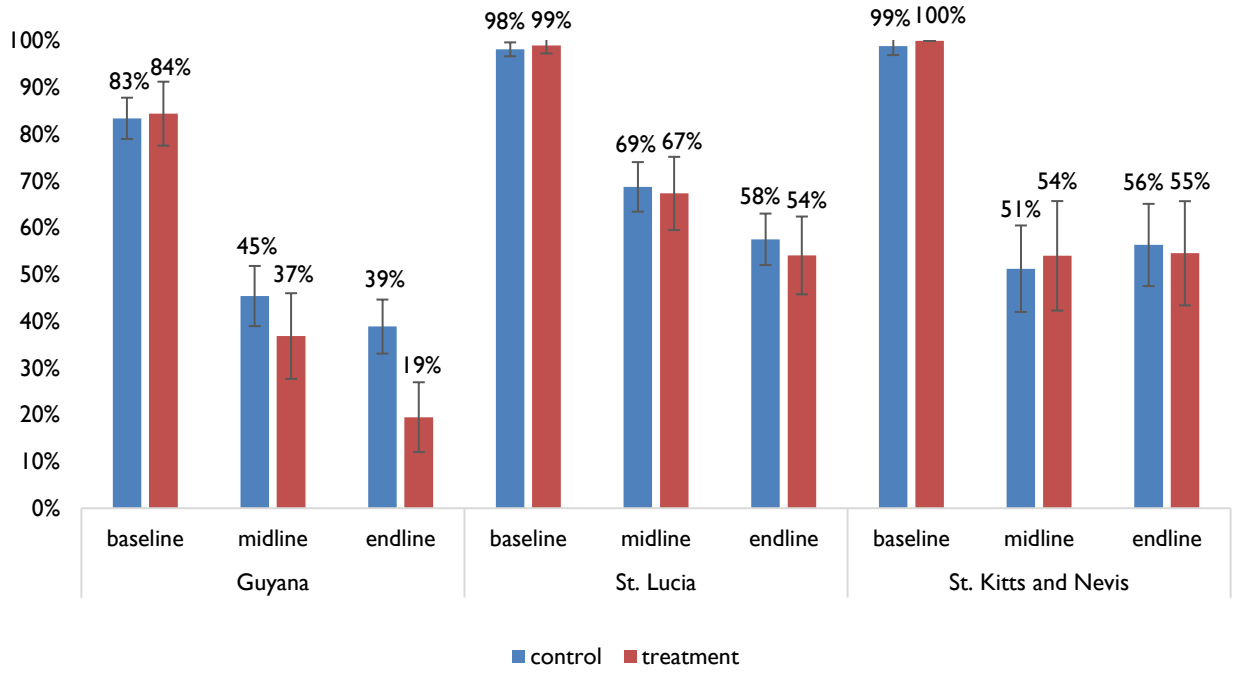


Figure 6 displays the mean number of risk factors by data collection period for the ITT and TOT samples. The TOT and ITT samples trend in the same direction and with similar change sizes from baseline to endline. As above, the most salient finding is the drop in both treatment and control youth; however, the "difference-in-difference" between treatment and control is statistically significant for both the TOT and ITT samples.

Figure 6: Average number of risk factors by data collection period, for TOT and ITT samples (with standard error bars)

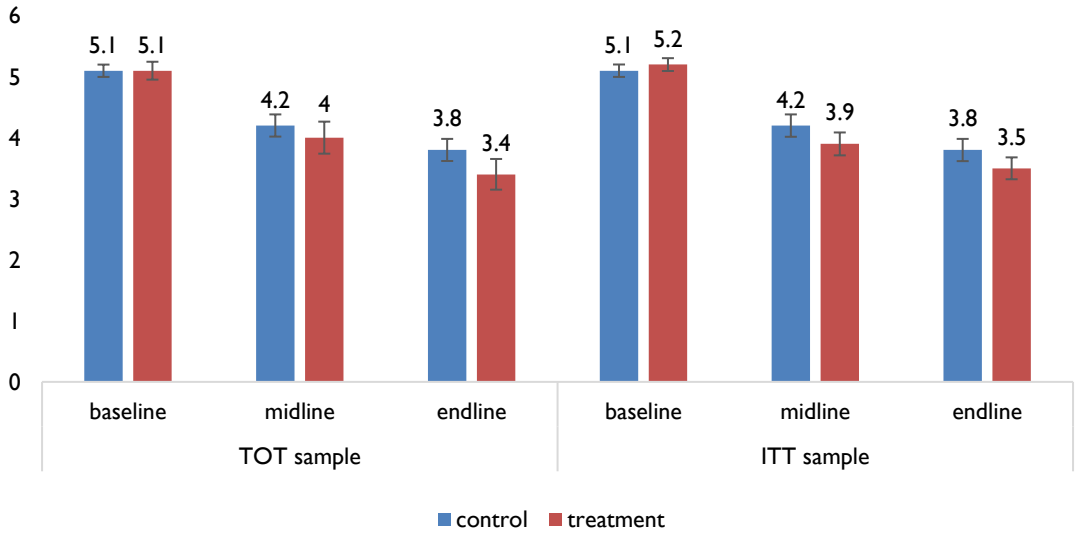


Figure 7 also compares the mean number of risk factors, but they are disaggregated by country and for the TOT sample only. Guyanese youth have the largest reduction in the mean number of risk factors, starting at 4.9 on average at baseline and dropping to 2.2 on average at endline for the treatment group. Therefore, in Guyana, the treatment group reduction in average scores is 2.7 points, compared with 1.7 points in the control group. This difference is statistically significant at the 95 percent confidence level. In St. Lucia, there is a downward trend among both treatment and control youth, but there is a smaller gap between treatment and control at endline compared to Guyana, and the difference is not statistically significant. In St. Kitts and Nevis, the treatment group experienced smaller reductions in average number of risk factors compared with the control group. Treatment youth in St. Kitts and Nevis reduced their average scores by 1.1 points from baseline to endline, while control youth in the same country reduced their average scores by 1.2 points. Again, these pooled analyses do not account for inter-community variation in effects, but more sophisticated models produce similar findings.

Figure 7: Average number of risk factors by data collection period for TOT sample, by country (with standard error bars)

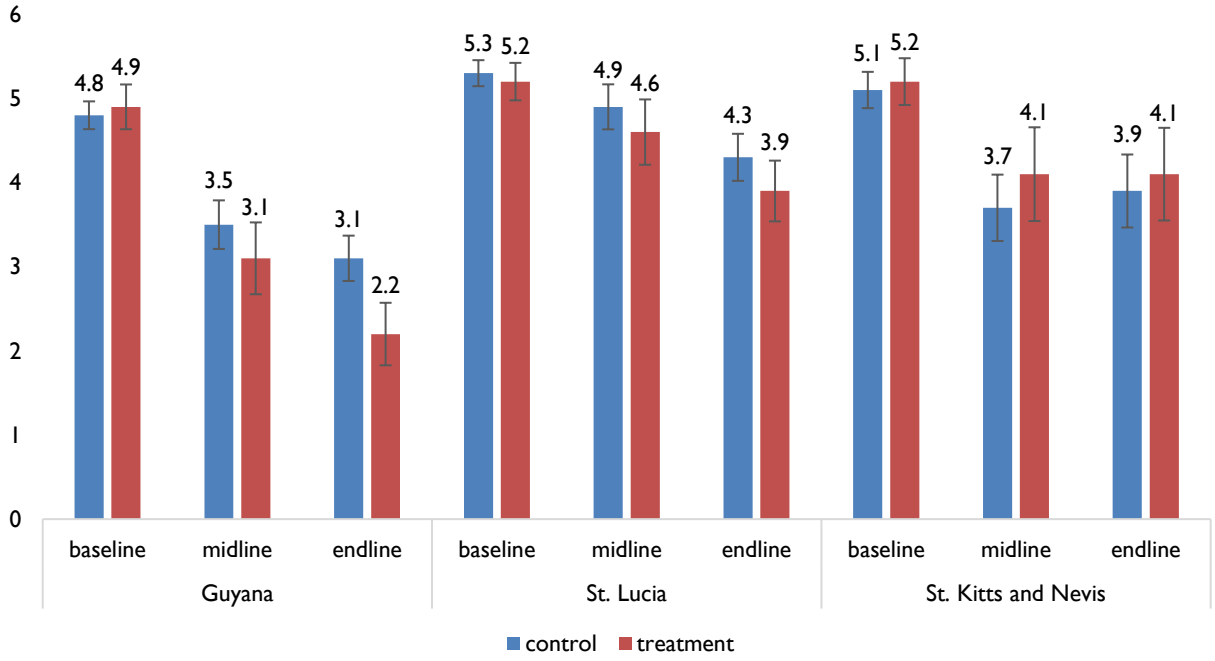


Figure 8 displays the share of youth that experienced an increase, decrease or no change in the number of risk factors from baseline to endline, disaggregated by treatment status. Most youth, regardless of treatment status, saw reductions in their overall risk factors. YSET scores decreased for 64 percent of control youth, 73 percent of youth who took up and completed treatment (TOT), and 71 percent of youth randomized to treatment (ITT).

Figure 8: Share of youth by risk level change from baseline to endline

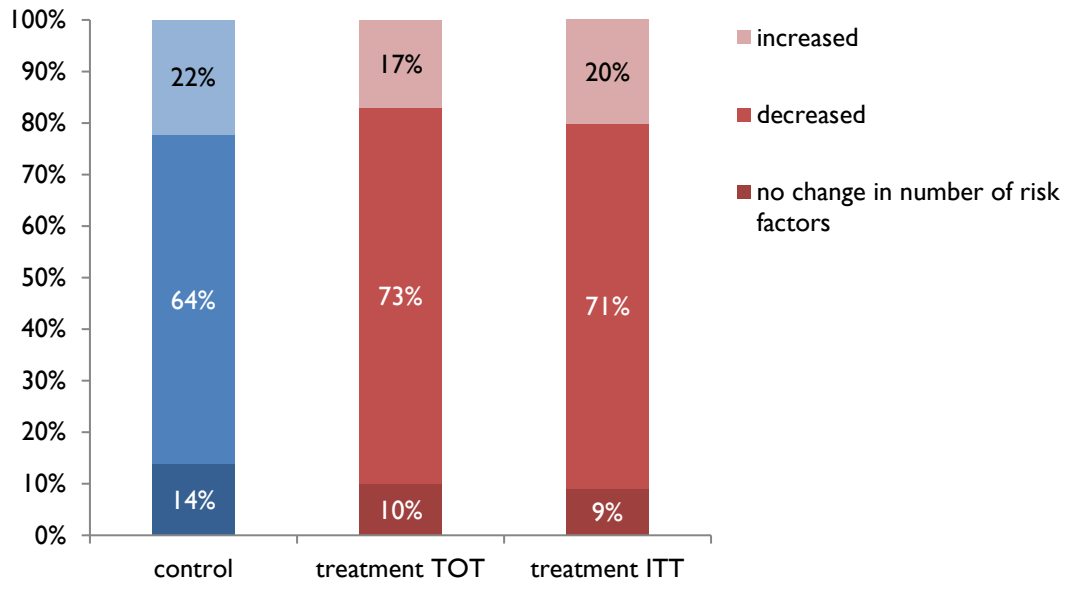
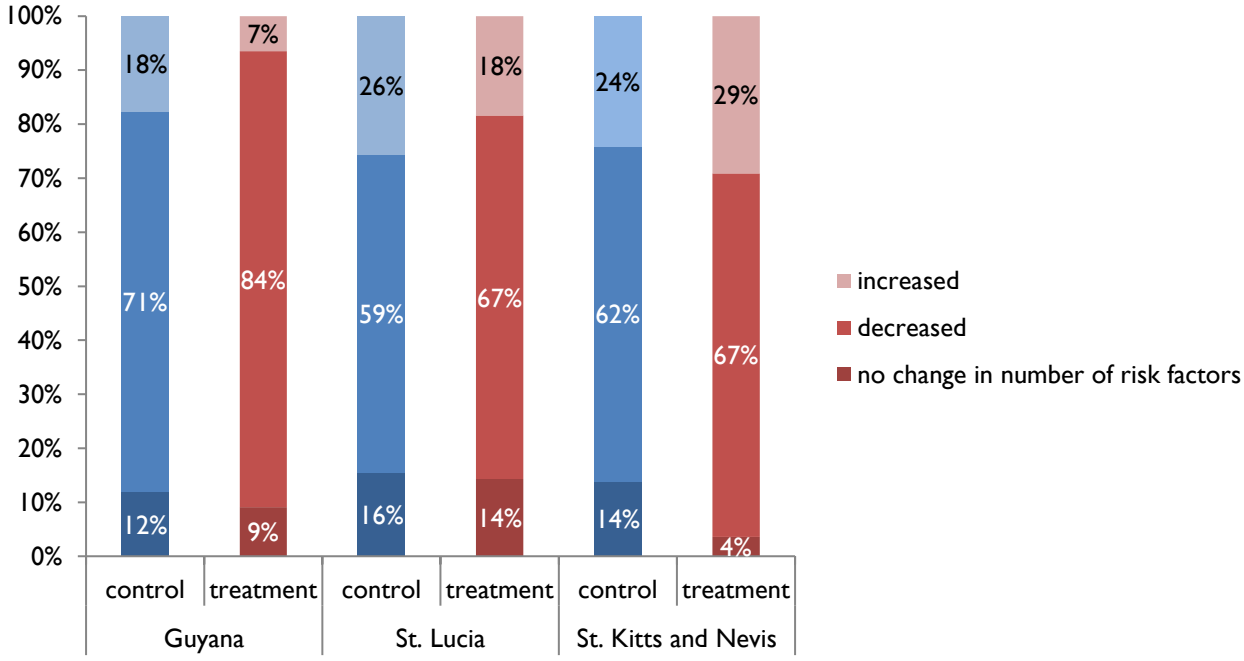


Figure 9 shows the same outcome of share of youth that experienced an increase, decrease or no change in the number of risk factors from baseline to endline, but disaggregated by country and for the TOT sample only. In Guyana, 84 percent of treatment youth saw a decrease in their number of risk factors, and only seven percent saw an increase in their number of risk factors. In St. Kitts and Nevis, however, a large share of treatment youth (29 percent) saw an increase in their number of risk factors, which is higher than the share of control youth in the St. Kitts and Nevis who experienced an increase.

Figure 9: Share of youth by risk level change from baseline to endline, by country, for the TOT sample



To a certain extent, the results by country are a bit surprising, because although the intervention was standardized across all three countries, the qualitative research trips at midline revealed what appeared to be a more intense approach taken by counselors in St. Lucia and St. Kitts and Nevis compared to Guyana. While we do not have an accurate measurement, qualitative research suggests that the frequency of meetings and extra communication (such as messages) between counselors and families appeared to be higher in St. Lucia than in the other countries. Many of the St. Lucian counselors reported having reached out to the youth's teachers or other school officials to be informed about how the youth are doing in school. In some instances, St. Lucian counselors were able to help the family acquire supplies for school or after-school activities, waive program fees, and prepared Christmas food baskets for them.

In St. Kitts and Nevis, we learned from CFYR that a lot of work was done initially to make sure families met basic needs before the intervention started. While this engagement was not required of them, counselors helped youth with enrolling in vocational school, filling out job applications and referring them to jobs, and helping families access the social assistance they need. Similar to counselors in St. Lucia, St. Kitts and Nevis counselors appeared to have more frequent communication and meetings with participants. Caregivers said they can contact their counselors anytime via WhatsApp or with a phone call. Although we heard in the other countries that counselors might travel to a community together, St. Kitts and Nevis appeared to be the only country where they co-counsel, meaning two counselors did the sessions together. The St. Kitts and Nevis counselors shared a common office, held weekly group meetings with their supervisors, and shared with each other in between group sessions as well. They were all very familiar with each other's cases. Despite these more intensive approaches taken in St. Lucia and St. Kitts and Nevis, it was youth in Guyana who experienced the largest drops in risk, as measured by the YSET.

DIFFERENCE-IN-DIFFERENCES MODELS

We ran four difference-in-differences models on each outcome of interest. The fifth model tests for heterogeneous effects by gender, age, country, school status and risk level (high/low). Table 5 includes a brief description of what each model does.

Table 5: Difference-in-difference models used for analysis

Model	Description
1. Pooled, simple difference-in-differences	Assumes all youth are drawn from the same population. Not adequate.
2. Random effects with clustered standard errors	Acknowledges that youth in a given community might be different on outcome variables than those in another community.
3. Random effects with clustered standard errors and community dummies	Acknowledges differences across communities while looking for differences within and between youth.
4. Fixed effects with clustered standard errors	Acknowledges differences across communities while looking for differences within youth.
5. Random effects with clustered standard errors and community dummies and interactions for age, gender, school status, risk level (high/low) and country	Acknowledges differences across communities while testing for heterogeneous effects. Statistical power is a concern in these models.

ITT ANALYSIS RESULTS

We examined the effect of the treatment for treatment youth in the ITT sample, which includes all youth randomized to treatment and control. The ITT sample is agnostic to participation status, meaning it includes youth regardless of whether they participated in the program or were approached by CFYR. The analysis ignores non-compliance, protocol deviations, withdrawal, and anything else that happens after randomization. ITT analysis results generally offer a more conservative estimate of program effectiveness, accepting that non-compliance and other deviations from the plan are likely to happen in real practice. The ITT analysis is valuable because its results indicate how effective an intervention is on the population of interest overall, and not just on those youth that were able to complete the program. The ability to reach and retain program participants until graduation is an important aspect of any program.

The treatment effect on the ITT sample is statistically significant for the "mean number of risk factors" and "four or more risk factors" outcomes, and for a few of the scales (the complete table and interpretation can be found in

ANNEX V: ADDITIONAL YSET SURVEY RESULTS). The coefficients are smaller than those for the TOT sample, which is not surprising given that only 48 percent of the ITT treatment sample actually completed the program.

TOT ANALYSIS RESULTS

Table 6 shows the means for the control and treatment groups on each of the outcome variables, at baseline, midline and endline. The youth included in the midline sample are the same that are included in the panel baseline-endline sample, minus the youth who could not be reached at midline. In the regressions, we compare only baseline and endline YSET scores, excluding midline.

We ran difference-in-differences regressions on each of the main outcomes using four different models (Table 5), and in the final column in the table below we report the coefficients and standard errors for the third model (with random effects, clustered standard errors and community dummies). The results with the other models are very similar to the ones reported here, both in terms of coefficient sizes and level of statistical significance.

The differences between control and treatment are statistically significant for the "mean number of risk factors" and "four or more risk factors" outcomes, and scales F and T. On average, treatment youth experienced a 34 percent reduction in their mean number of risk factors from baseline to endline (from 5.12 to 3.37), and control youth experienced a 26 percent reduction (from 5.09 to 3.75). The coefficient for the "mean number of risk factors" outcome tells us treatment youth experienced a reduction that is 0.418 points larger (in the YSET scale) than that of the control group. The reduction in the share of treatment youth who are "eligible" (four or more risk factors) was 9.3 percent larger than the reduction for the control group. In the F scale, the share of treatment youth who reduced risk in this scale was nine percent greater than that of the control group. In the T scale, the share of treatment youth who reduced risk in this scale was 10 percent greater than that of the control group.

Table 6: Difference-in-differences for the second, third and fourth set of outcomes for the TOT sample

	Means						Coef.
	Baseline		Midline		Endline		B-E
	C (n=499)	**T (n=230)**	**C (n=451)**	**T (n=224)**	**C (n=499)**	**T (n=230)**	**Diff-in-diff**
Mean Number of Risk Factors	5.09	5.12	4.16	3.97	3.75	3.37	**-0.418** (0.19)**
Four or More Risk Factors	0.93	0.94	0.57	0.54	0.50	0.43	**-0.093** (0.04)**
B: Weak Parental Monitoring	0.45	0.43	0.41	0.41	0.44	0.36	-0.070 (0.05)
C: Critical Life Events	0.61	0.62	0.45	0.47	0.39	0.40	0.005 (0.05)
DE: Impulsive Risk Taking	0.75	0.81	0.57	0.54	0.47	0.46	-0.069 (0.05)
EG: Risky Behaviors	0.37	0.37	0.27	0.28	0.22	0.18	-0.041 (0.04)
F: Guilt Neutralization	0.81	0.78	0.66	0.62	0.63	0.51	**-0.090* (0.05)**
G: Negative Peer Influence	0.39	0.37	0.38	0.34	0.35	0.30	-0.025 (0.05)
H: Peer Delinquency	0.49	0.44	0.39	0.37	0.34	0.30	0.007 (0.05)
IJ: Self-Reported Delinquency	0.74	0.74	0.57	0.53	0.48	0.45	-0.037 (0.05)
T: Family Antisocial Influence	0.48	0.56	0.46	0.41	0.43	0.41	**-0.098** (0.05)**

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 7 shows the results for another set of outcomes: average scores on each scale. Scale F has strong results, with statistically significant results both in the average scores in

Table 7, and share of youth at risk in Table 6. The reduction in average scores in scale F was 0.112 points larger for the treatment group than the control group. The other scale that has statistically significant results for average scores is DE, in which the reduction in average scores was 0.16 points larger for the treatment group than the control group.

Table 7: Difference-in-differences for the first set of outcomes for the TOT sample

	Means						Coef.
	Baseline		Midline		Endline		B-E
	C (n=499)	T (n=230)	C (n=450)	T (n=224)	C (n=499)	T (n=230)	Diff-in-diff
A: Anti-Social Tendencies (Mean)	2.38	2.41	2.25	2.21	2.15	2.15	-0.020 (0.05)
B: Weak Parental Monitoring (Mean)	2.15	2.12	2.06	2.02	2.14	1.98	-0.121 (0.08)
C: Critical Life Events (Count)	3.71	3.87	3.27	3.37	3.06	3.13	-0.093 (0.16)
DE: Impulsive Risk Taking (Mean)	3.33	3.43	3.05	3.02	2.91	2.85	**-0.160** (0.08)**
EG: Risky Behaviors (Count)	2.10	2.03	1.63	1.59	1.35	1.12	-0.166 (0.17)
F: Guilt Neutralization (Mean)	2.94	2.92	2.76	2.72	2.75	2.62	**-0.112** (0.06)*
FSH: Horizontal Family (Mean)	2.35	2.27	2.28	2.27	2.30	2.22	-0.011 (0.06)
FSV: Vertical Family (Mean)	2.57	2.48	2.59	2.58	2.54	2.54	0.089 (0.08)
G: Negative Peer Influence (Mean)	1.95	1.87	1.96	1.84	1.91	1.75	-0.074 (0.08)
H: Peer Delinquency (Mean)	1.90	1.86	1.73	1.70	1.69	1.66	0.002 (0.06)
IJ: Self-Reported Delinquency (Count)	4.51	4.57	4.08	3.79	3.50	3.20	-0.368 (0.31)
T: Family Antisocial Influence (Sum)	5.89	5.90	5.81	5.60	5.76	5.53	-0.254 (0.24)

Note: *** p<0.01, ** p<0.05, * p<0.1

The following charts include the questions from the two scales that had statistically significant results in their average score changes: DE and F. In Scale DE, youth are asked questions about impulsive risk taking, and how strongly they agree or disagree with each statement. In Figure 10 and Figure 11 we can see that, while both the treatment and control group show improvements between baseline and endline, the treatment group improved more, particularly in the share of youth selecting the answer "strongly disagree." Also, the share of treatment youth selecting "strongly agree" always goes down, while for control youth it stayed the same in the second question and went up in the first question.

Figure 10: Scale DE: Youth preference for impulsive risk taking

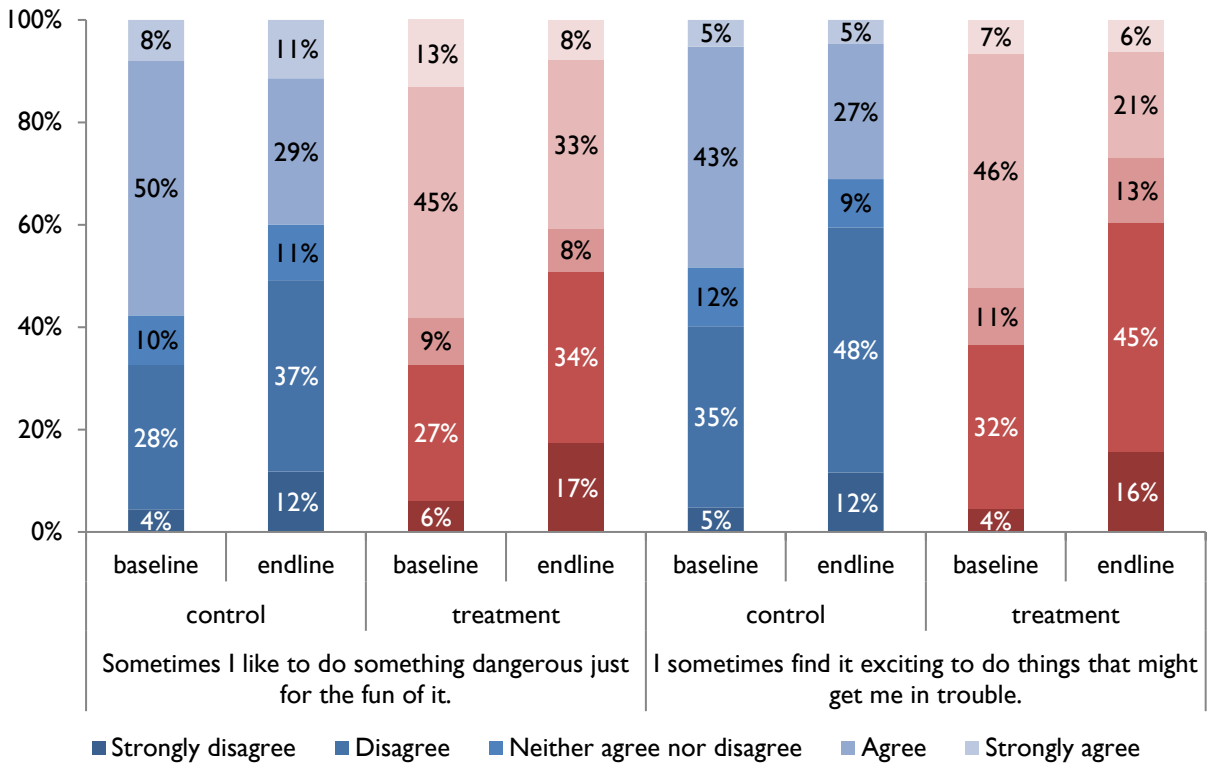
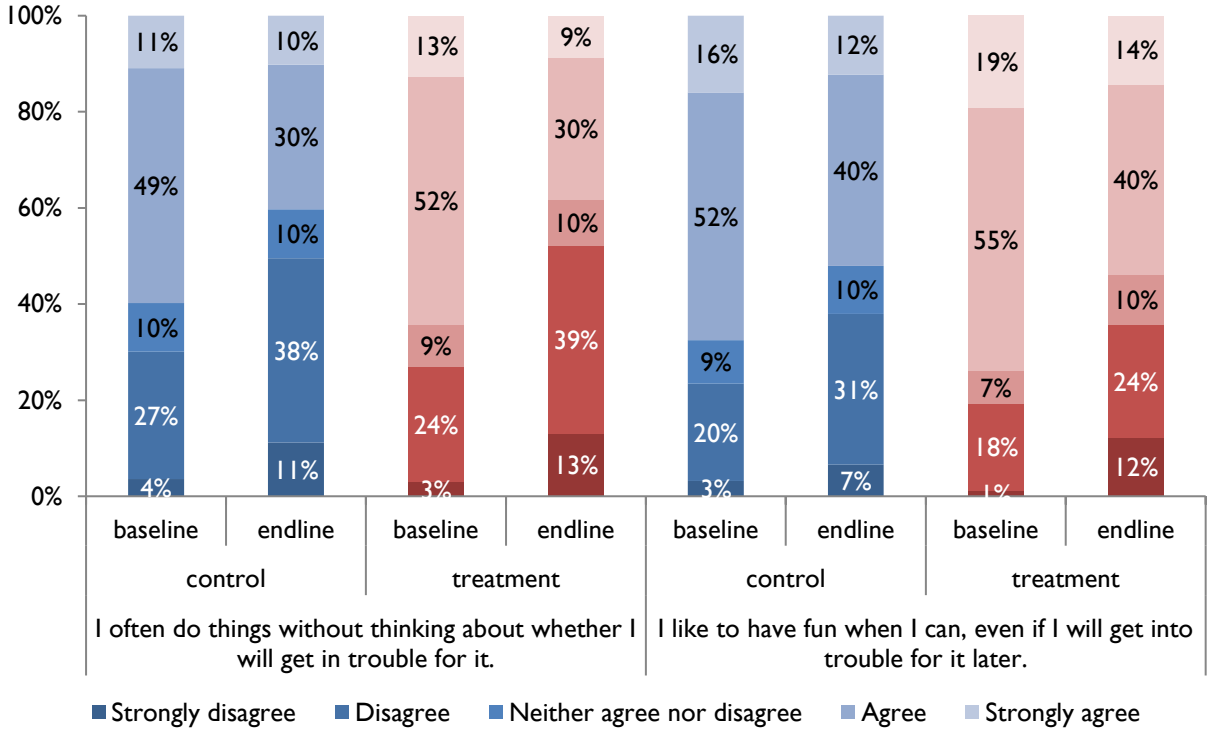


Figure 11: Scale DE: Youth comfortability with negative consequences of impulsive risk tasking



In Scale F, youth are prompted with statements about guilt neutralization and asked whether they agree or disagree with that statement. The statements involve lying, stealing and using physical violence. Again, we see that the share of youth selecting "strongly disagree" increased more for the treatment group than the control group. Also, we see that more youth in general are comfortable with lying (Figure 12) or using physical violence (Figure 14), than they are comfortable with stealing (Figure 13). Much smaller shares of youth in general agree that it is okay to steal from somebody who is rich or to steal from a store.

Figure 12: Scale F: Youth comfortability with lying

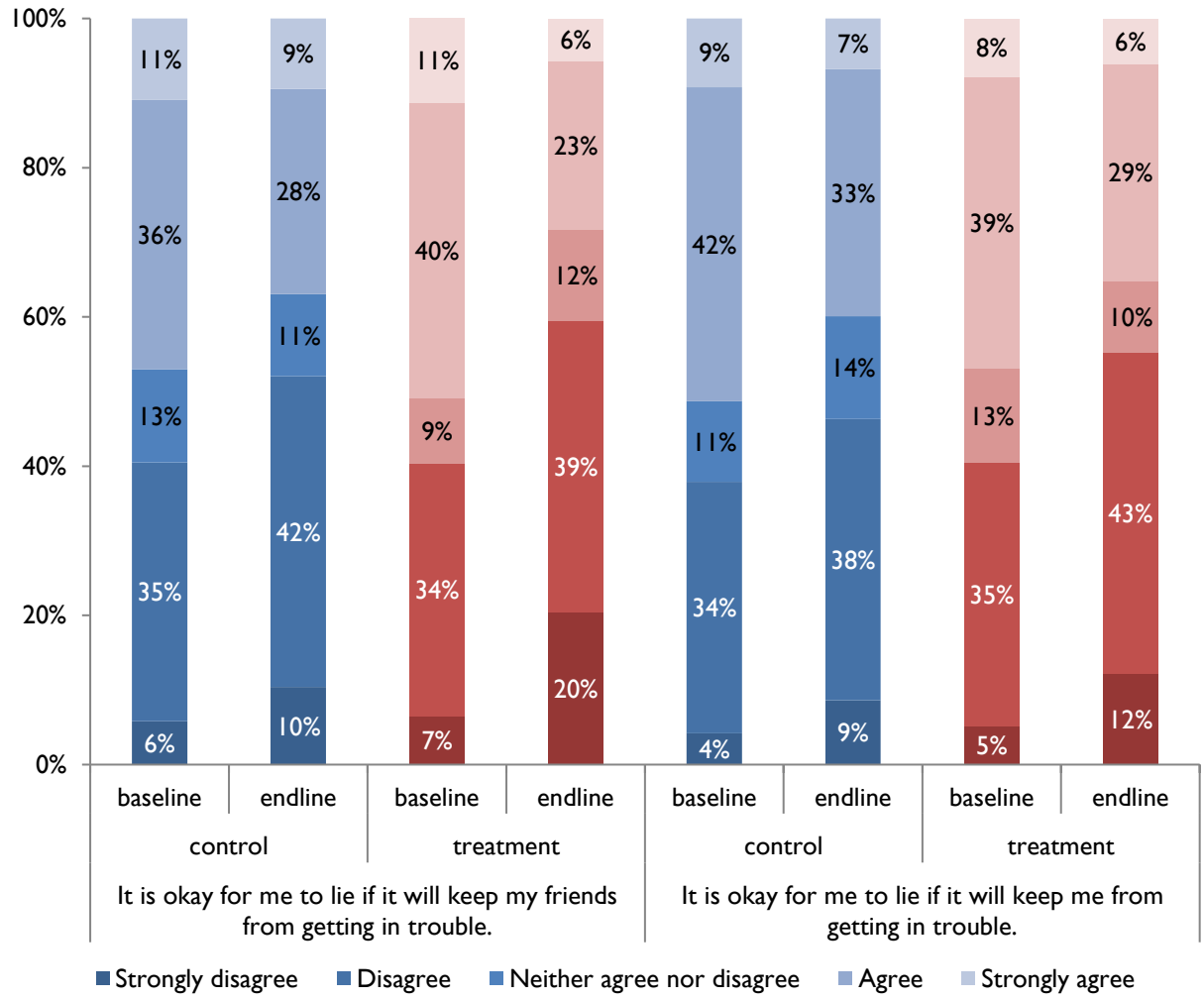


Figure 13: Scale F: Youth comfortability with stealing

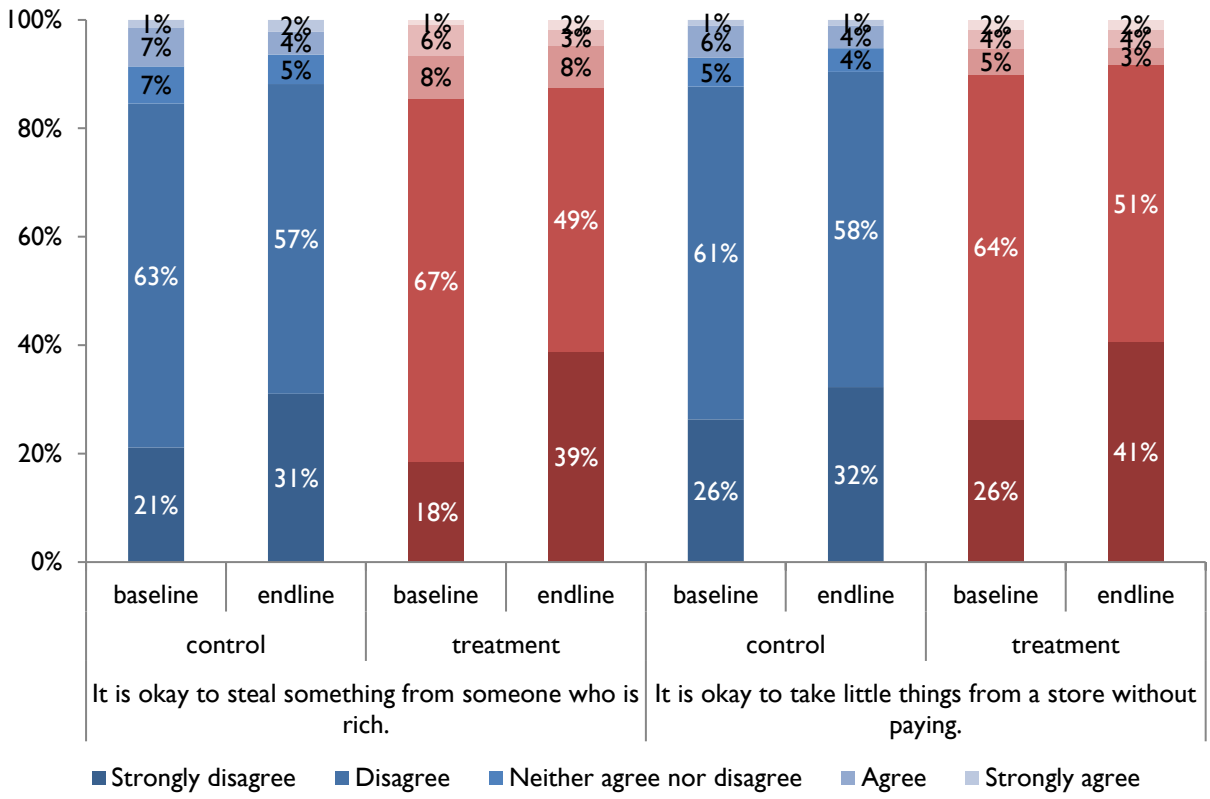
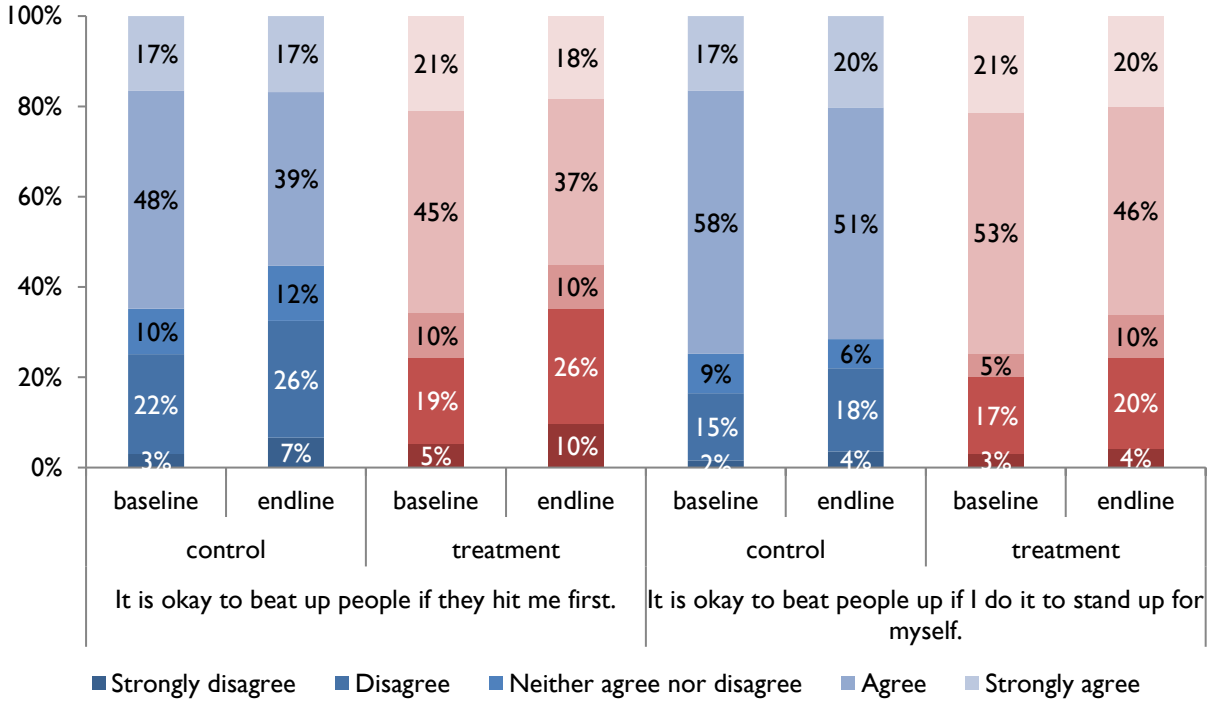


Figure 14: Scale F: Youth comfortability with using physical violence



TOT ANALYSIS RESULTS WITH MATCHING

To account for the imbalance found in the TOT sample (see ANNEX III: BALANCE TESTS for the balance tables), we also ran the TOT analysis on a matched sample of treatment and control youth. More details on the matching technique used can be found in the Methodology section. The results are less positive overall for the program with the matched sample, with no statistically significant differences between treatment and control for the "mean number of risk factors" and "four or more risk factors" outcomes. Nevertheless, Scales F and DE retained their level of statistical significance and got slightly larger coefficients in the matched TOT sample. The full results are located in

ANNEX V: ADDITIONAL YSET SURVEY RESULTS.

Null results can occur due to an evaluation being underpowered, so we calculated the ex-post MDE on all YSET outcomes. The MDE tell us what the smallest effect is, in terms of standard deviations of each outcome, that was detectable considering the level of power and statistical significance chosen. With 80 percent power at a significance level of 0.05, we used the following formula to calculate MDE:[14]

$$MDE = 2.8 \times SE(\hat{\beta})$$

The ex-post MDE of the matched TOT and TOT samples contribute to the interpretation of the results, especially the null results. The MDE provides a way to distinguish between null effects that may be associated with small changes and null effects that may point to an indeterminate relationship. Table 8 shows the control group mean at baseline for each of the outcomes, along with the MDE for that outcome, for the two TOT samples. For the outcomes "mean number of risk factors" and "four or more risk factors," we have enough power in both samples to determine whether the coefficients we estimated (see Table 34 and Table 35 in

[14]For an example of an academic paper that calculates ex-post MDE the same way and for the same reason, see Haushofer and Shapiro, 2016.

ANNEX V: ADDITIONAL YSET SURVEY RESULTS) were statistically significant. In other words, the lack of statistical significance in the matched sample for those two outcomes is not the result of a lack of power, and we can be confident that there is no significant difference between treatment and a matched control group. For some of the scales, the coefficients are smaller than the MDE, but those coefficients are quite small (meaning that they indicate a very small difference between treatment and control). Our conclusion is that we should not be concerned about power in these results, including in the matched sample for the main outcomes of "mean number of risk factors" and "four or more risk factors."

Table 8: Ex-post MDE for YSET indicators

	Matched TOT		TOT	
	Control Mean at Baseline (n=301)	MDE	Control Mean at Baseline (n=499)	MDE
Mean Number	5.11	0.223	5.09	0.173
Four or More Risk Factors	0.94	0.042	0.93	0.033
A: Anti-Social Tendencies (Mean)	2.35	0.084	2.38	0.065
B: Weak Parental Monitoring (Mean)	2.12	0.126	2.15	0.098
C: Critical Life Events (Count)	3.75	0.219	3.71	0.170
DE: Impulsive Risk Tasking (Mean)	3.35	0.108	3.33	0.084
EG: Risky Behaviors (Count)	2.00	0.311	2.10	0.242
F: Guilt Neutralization (Mean)	2.92	0.087	2.94	0.068
FSH: Horizontal Family (Mean)	2.33	0.100	2.35	0.078
FSV: Vertical Family (Mean)	2.57	0.137	2.57	0.107
G: Negative Peer Influence (Mean)	1.96	0.145	1.95	0.113
H: Peer Delinquency (Mean)	1.88	0.102	1.90	0.079
IJ: Self-Reported Delinquency (Count)	4.35	0.515	4.51	0.400
T: Family Antisocial Influence (Sum)	6.04	0.415	5.89	0.322

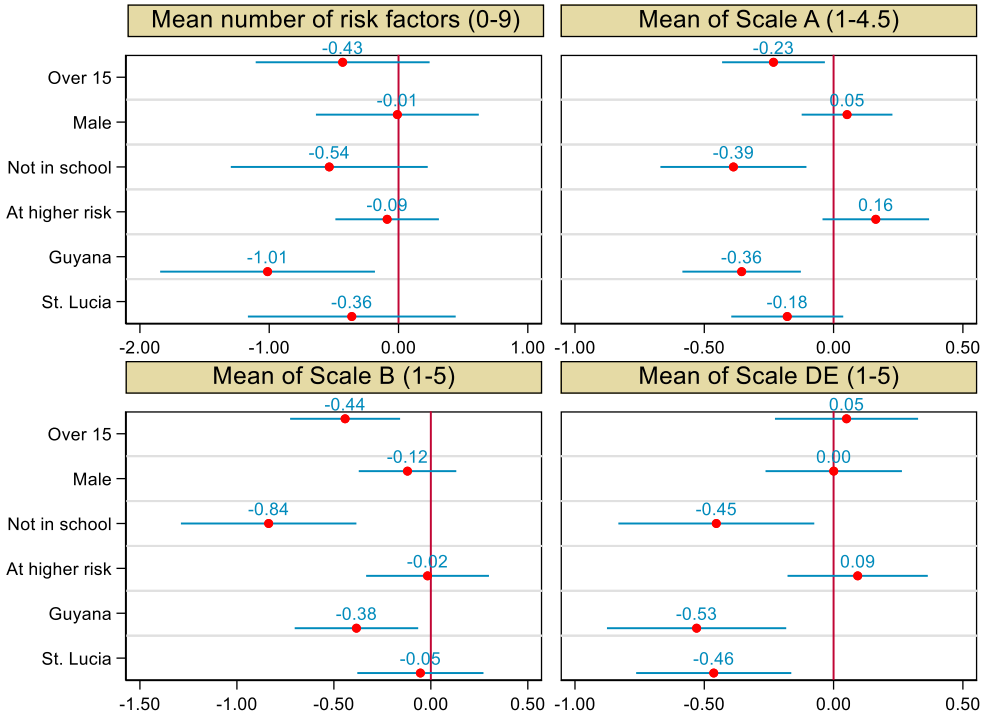
HETEROGENEOUS TREATMENT EFFECTS

We ran a heterogeneous treatment effect analysis on the full TOT sample to check if there are statistically significant differences in risk reduction by group. Specifically, we explored if the intervention had different effects based on: (1) age (14 and under, or over 15 years of age), (2) gender (male or female), (3) school status (in or out of school at the time of the survey), (4) risk level (higher risk at six or more risk factors in the YSET modules, or lower risk at five or fewer risk factors), and (5) country. We ran the analysis at the 90 percent confidence level and with the preferred model of random effects, clustered standard errors and community dummies. We should note that the sample sizes of these subgroups are only large enough to confidently detect large heterogeneous effects.

Figure 15 provides estimates of the heterogeneous effects of these five factors using the full TOT sample, which we have used because of the greater number of observations. The numbers presented in the figure are based on several statistical models, each testing the interaction between treatment and the selected variable. In the case of country, for example, the model tells us the extent to which the intervention has a different effect in Guyana or St. Lucia as compared with St. Kitts and Nevis, which is not pictured in the figure as it is the point of comparison. Just like for the main results, negative coefficients indicate a greater reduction in risk for that group vis-a-vis its comparison group. The "Mean number of risk factors" figure illustrates that treatment in Guyana has a one-point reduction in risk on a zero to nine scale compared with St. Kitts and Nevis, and this difference is statistically significant. (We should note, however, that this difference is not statistically significant with the matched data.) The other coefficients that show greater

reductions in risk are: 1) the mean score in the A and B scales for the over 15 years of age group, the out-of-school youth, and Guyanese youth, in comparison with youth from St. Kitts and Nevis; 2) for the mean score in the DE scale for and Guyanese and St. Lucian youth in comparison with St. Kitts and Nevis.

Figure 15: Differences in risk reduction by group using the full TOT sample



Out-of-school youth experienced larger treatment effects than youth in school for the mean scores in scales A, B, DE and FSV (the last one is reported in

ANNEX V: ADDITIONAL YSET SURVEY RESULTS). These include three out of four scales that the intervention focused on during the first six months (namely, A, B, FSH and FSV). These results suggest that the intervention may have been more effective for out-school-youth. However, we also note that a larger share of out-of-school youth are present in Guyana compared to the other countries, particularly in comparison with St. Kitts and Nevis. So it may be that a more effective intervention in Guyana helped the out-of-school youth coefficients become larger and statistically significant, or instead it may be that the intervention is more effective for out-of-school youth and that helped the treatment coefficients become larger and statistically significant for Guyana. Our sample is not large enough to look at differences between youth that are in or out of school in each of the countries, to help us decide whether one of these caused the other.

These were the outcome variables where we saw the largest number of statistically significant differences, and the most interesting results. In

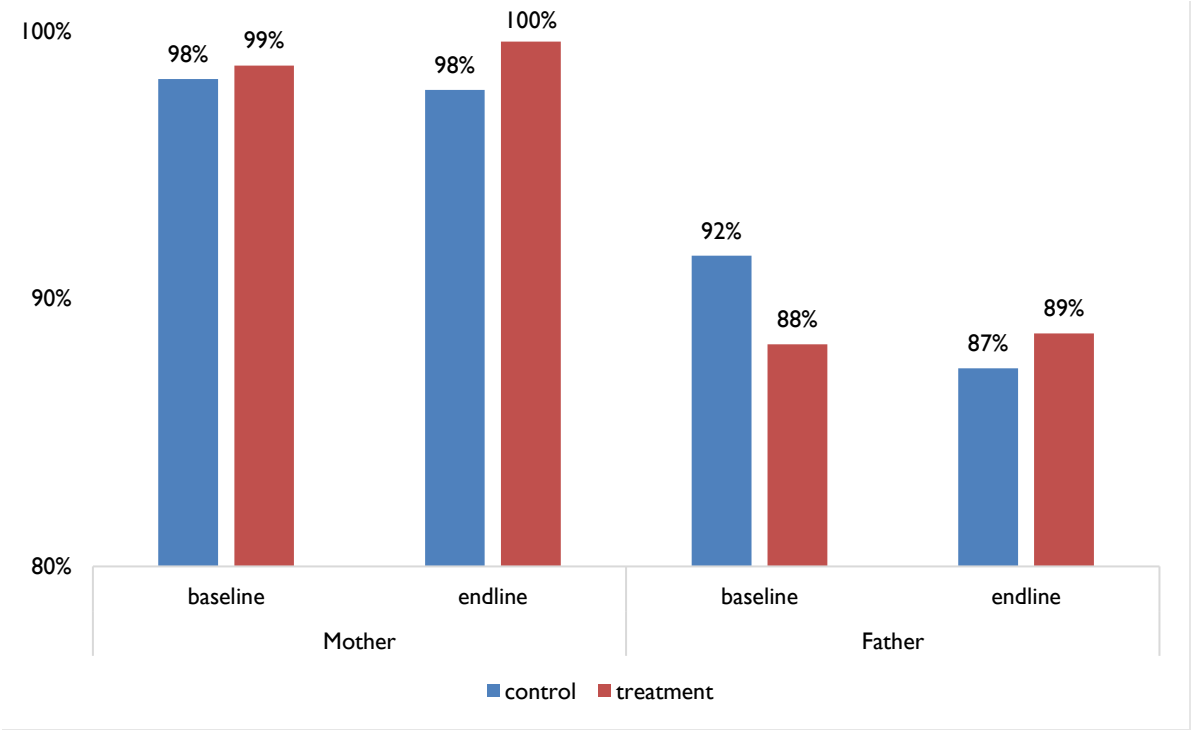
ANNEX V: ADDITIONAL YSET SURVEY RESULTS, we also include a chart with statistically significant results in four other outcome variables.

MOTHER AND FATHER MODULES

During qualitative fieldwork, the SI researchers heard from family counselors and youth that there was an effort being made on the part of counselors to reach out to fathers and include them in the counseling. Some treatment youth then reported to us that their previously absent fathers were getting more involved. Given that treatment and control youth were asked specific questions about their relationship with their mother/stepmother/female guardian, and their father/stepfather/male guardian, we decided to examine this quantitatively.

Using the full TOT sample, we found that at both baseline and endline, almost all youth reported having a mother/stepmother/female guardian. We do, however, observe a slight increase in the percent of treatment youth who report the presence of a father/stepfather/male guardian, while the percentage of control youth who reported having a father/stepfather/male guardian decreased from 92 percent at baseline to 87 percent at endline. Nonetheless, this difference-in-differences was not found to be statistically significant at the 90 percent confidence level.

Figure 16: Share of youth reporting they have a mother/stepmother/female guardian, and a father/stepfather/male guardian (n=729)



Youth who responded "yes" to the previous question were then asked about their relationship with said caregiver. There are no visible differences between control and treatment regarding their relationships with their mother (see ANNEX VIII: ADDITIONAL CHARTS).

For the questions concerning fathers, however, there is one interesting difference between treatment and control youth. While the share of control youth answering "Never" to the following questions remains

the same or even increases in some instances, the share of treatment youth who select "Never" decreases over time for every question. The overall trend for the "Always" and "Often" answers is similar for the treatment and control groups over time.

Figure 17: Questions concerning the relationship with father/stepfather/male guardian (n=660)

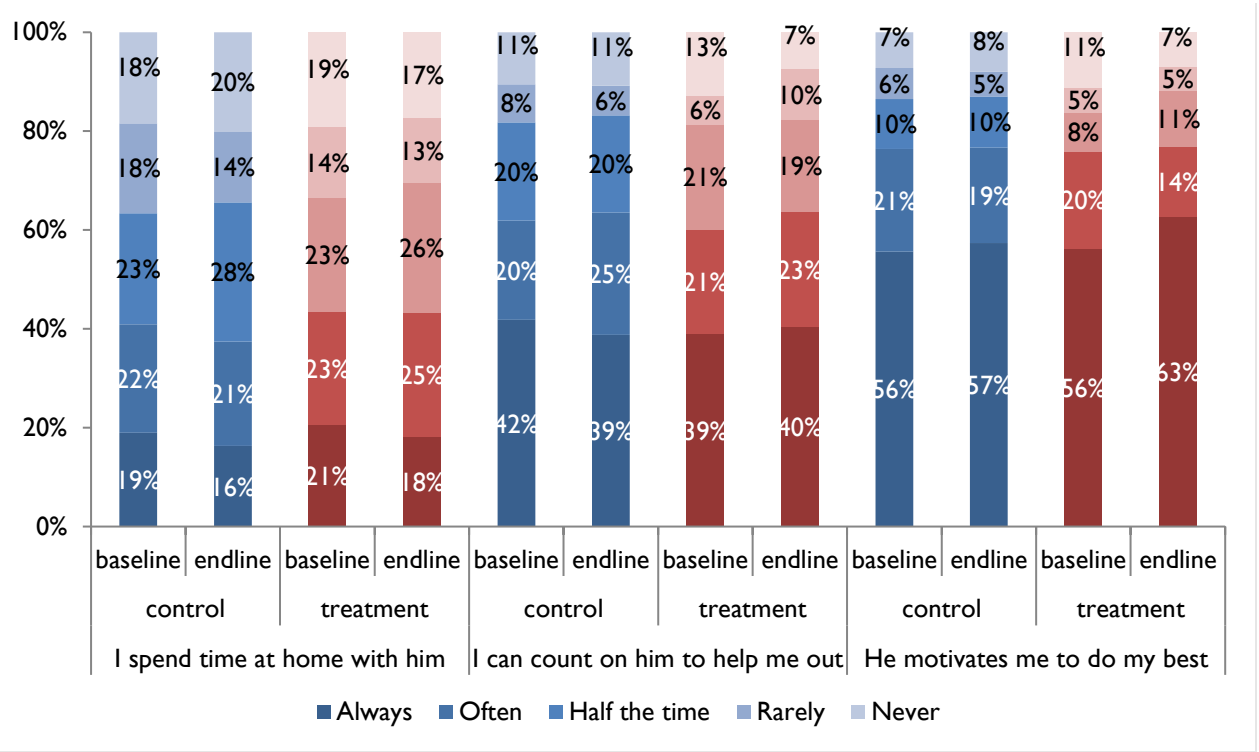
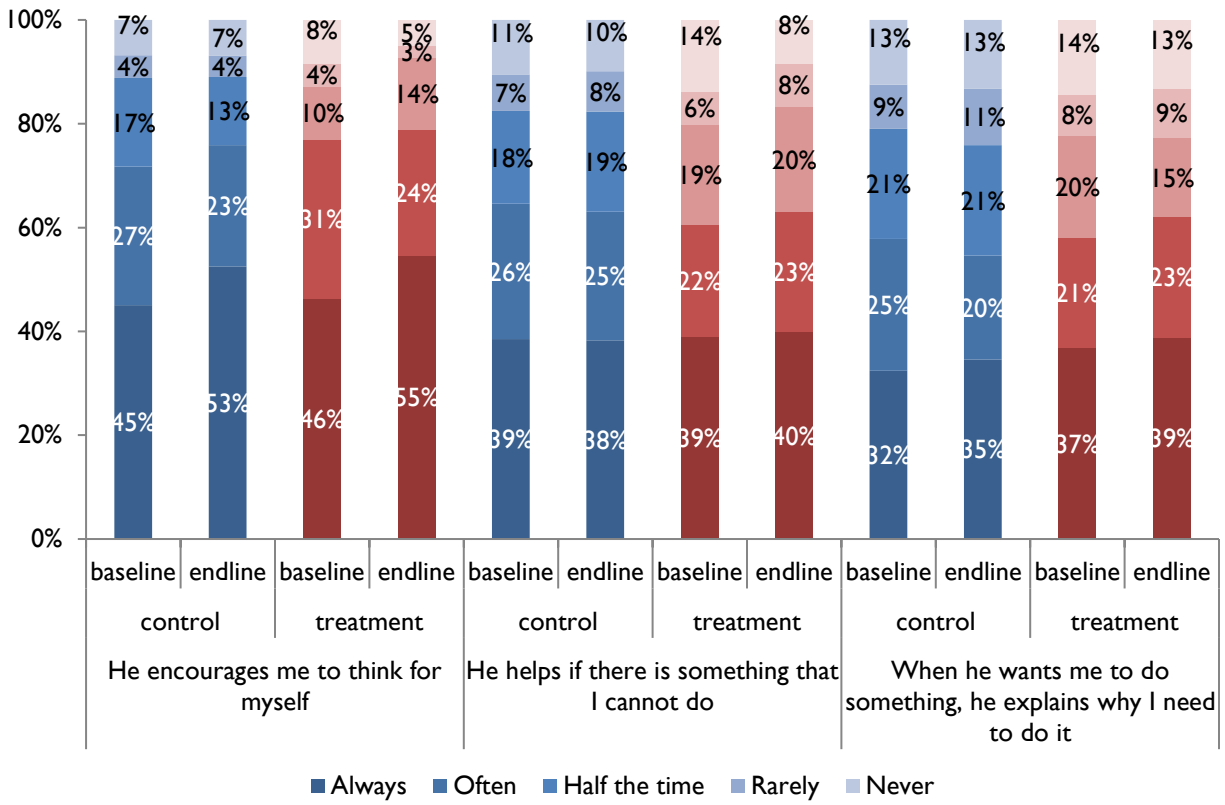


Figure 18: Questions concerning the relationship with father/stepfather/male guardian (n=660)



FINDINGS BASED ON THE CAREGIVER SURVEY

The CFYR *Family Matters* program aims to help families better protect and guide youth by providing structured family counseling to youth and their families. Both caregivers and youth regularly meet with counselors and are provided with tools to improve family dynamics and youth behavior in order to effect change within the household. Through the program, family members are expected to interact more frequently, communicate more effectively, make joint decisions more regularly, demonstrate a stronger sense of family pride, and exhibit a stronger reliance on community networks for support. While the program's ultimate goal is to improve youth behavior and outcomes, the pathway for change is through both the youth and his/her family members. As such, we examine the impact of the program on not only youth but on their caregivers as well. This section defines the caregiver outcomes of interest and explores the main results of the caregiver analysis, as well as ancillary results including heterogeneous treatment effects, an assessment of family dynamics and family type risk determinant analysis.

OUTCOMES OF INTEREST

The caregiver surveys for baseline and endline can be found in the supplementary Annex IX. The caregiver survey is a tool uniquely created for this evaluation. The questions draw from the YSET modules and/or capture outcomes expected to be generated by the program. The instrument was revised, pre-tested and piloted prior to implementation. Additional cleaning and internal checks were conducted to review questions to determine if any were problematic.

The survey questions are divided into modules focused on specific topics, and we used multiple questions on a given topic to construct a series of indices that could have potentially been impacted by the intervention. For ease of interpretation, in most cases the outcomes are simple additive indices, in which questions with similar answer scales were added together. In one case, we used a factor analysis (i.e., for a scale based on select questions from the Family cohesion index), and we used an average to calculate an index about how informed caregivers are of youth behavior and relationships. Caregiver outcomes of interest are:

1. **Extended family cohesion index**: Sum of Likert scale variables of how often youth visits any family members, calls any family members, and how often family members share family traditions.
2. **Family cohesion index**: Factor of Likert scale variables of the extent to which the respondent agrees that family members are involved in each other's lives, family members feel very close, family members are supportive during difficult times, family members can calmly discuss problems with each other, when angry, family members seldom say negative things about each other; and of Likert scale variables of the extent to which respondents are satisfied with the quality of communication between family members, family's ability to resolve conflicts and the amount of time spent together as a family.
3. **Family consensus index**: Sum of Likert scale variables of how often caregiver disagrees on how to raise or discipline youth; how often caregiver talks badly about other caregiver; and how often caregiver argues or fights in front of youth.

4. **Parenting locus of control index**: Sum of Likert scale variables of caregiver agreeing that children will get into trouble no matter what their parents do; the conditions in this community prevent caregiver from keeping his/her children out of trouble; parents can have a major influence on their kids; and it is primarily the caregiver's responsibility to keep his/her children out of trouble.
5. **Improvement self-evaluation**: Sum of Likert scale variables of caregiver agreeing that in the last 12 months, he/she has improved as a parent and in the last 12 months, his/her influence on his/her children has increased.
6. **Caregiver presence index**: Sum of continuous variables of the number of days per week the caregiver sits down to eat together with child; the number of days per week the caregiver helps child with schoolwork; the number of days per week the caregiver does something fun as a family with the child; and the number of days per week the caregiver does something religious as a family with the child.
7. **Informed parent index**: Average of Likert scale variables of how often child tells parent when he/she goes out, how often caregiver knows where child is when child is not at home or school, how often the caregiver knows who child is with when child is not at home or school, how often the child lets another adult know where he/she is if he/she does not let the caregiver know where he/she is.
8. **Parental authority index**: Sum of Likert scale variables of how often there are consequences if family rules are broken, how often the child listens to the caregiver when the caregiver has something important to say, how often the caregiver listens to the child when the child has something important to say, and how often the child and caregiver are able to talk and solve their problems, how often the caregiver has other adults he/she can count on to help raise the child and how often the caregiver is proud of the child.
9. **Youth behavior and relationships index**: Sum of when caregivers answer yes to whether they believe their child inhaled (sniffed, smoked or drank) drugs or substances that make you "high", skipped classes at school without an excuse or permission, lied about his/her age to get into some place or to buy something, purposely damaged or destroyed property that did not belong to him/her, carried a hidden weapon for protection, stolen or tried to steal something valuable, broken into a building to steal something, hit someone with the purpose of hurting him/her, attacked someone with a weapon, used a weapon or force to get money or things from people, been involved in group fights, been involved in gang fights, sold marijuana and illegal drugs.

For the extended family cohesion index, family adaptability and cohesion scale, family consensus index, parenting locus of control index, improvement self-evaluation, caregiver presence index, informed parent index, and parental authority index, the variables are reverse coded so that higher, positive values represent improved caregiver outcomes. For the youth behavior and relationships index, lower, negative values represent improved outcomes.

MAIN RESULTS OF THE TOT ANALYSIS

In accordance with the pre-analysis plan and to account for a slightly unbalanced sample, we run the analysis on the ITT, TOT, and matched TOT samples. We run the analysis on four models for each sample including a simple difference-in-differences regression, a random effects model with clustered standard errors, a random effects model with clustered standard errors and community dummies, and a fixed

effects model with clustered standard errors. The models used to analyze the caregiver survey are the same as the models used to analyze the YSET (see Table 5). We find minimal differences across samples and models and include results of the simple difference-in-differences model and random effects with clustered standard errors model in ANNEX VII: ADDITIONAL CAREGIVER SURVEY RESULTS.

The primary effects of the *Family Matters* program on caregiver outcomes are displayed in Table 9. The table compares treatment and control households in the TOT caregiver sample, so the treatment group includes all caregivers of youth who were randomized to treatment, were offered, took up and completed the program. Column 1 reports control means at baseline. Column 2 reports treatment means at baseline. Column 3 reports control means at endline. Column 4 reports treatment means at endline. Column 5 reports the treatment effect and standard error from the difference-in-differences regression with random effects, clustered standard errors and community dummies, which is our preferred model.

Unlike the YSET, we observe very little change between baseline and endline for either the treatment or the control caregivers. The effects of the program on caregiver indicators are small and not statistically significant. The model yields virtually the same results as the other three models we ran, as seen in ANNEX VII: ADDITIONAL CAREGIVER SURVEY RESULTS. Directionally, most results are not moving in the hypothesized direction; however, in all cases the indicators are close to zero.

Table 9: Difference-in-differences for the first set of outcomes, including false positives, for TOT

	Means				Coef.
	Baseline		Endline		B-E
	C (n=470)	T (n=225)	C (n=470)	T (n=225)	Diff-in-diff
Extended Family Cohesion Index	9	9.3	8.99	9.34	-0.073 (0.29)
Family Cohesion Index	-0.04	-0.03	-0.04	-0.03	-0.060 (0.09)
Family Consensus Index	5.7	6	5.7	5.9	-0.091 (0.32)
Parenting Locus of Control Index	13.5	13.5	13.4	13.3	-0.096 (0.20)
Improvement Self-Evaluation	8.3	8.5	8.4	8.6	0.014 (0.13)
Caregiver Presence Index	11	11.6	10.3	10.5	-0.425 (0.59)
Informed Parent Index	4	4	4.1	4	-0.142 (0.09)
Parental Authority Index	21.7	22.1	21	21.1	-0.331 (0.38)
Youth Behavior and Relationships Index	0.7	0.9	0.5	0.6	0.013 (0.12)

Given these results, there is little evidence that the *Family Matters* program meaningfully improves family dynamics or caregiver experiences with youth over the course of the year-long program. Caregivers randomized to treatment in the *Family Matters* program have similar experiences with their children as those caregivers who were randomized to control. Some caregiver indicators show discrete improvements while others show declines, and no indicators show statistically significant results. Thus,

the empirical evidence fails to prove that caregivers' management of their households and experiences with youth are improved by the *Family Matters* program.

Treatment caregivers may decline or remain largely unchanged for a number of reasons from baseline to endline, in contrast with the aim of the intervention. Two complementary theories emerged from our qualitative research to explain this phenomenon. First, caregivers may not be aware of or admit problems at baseline. By endline, however, treatment caregivers have been given the tools to identify and verbalize problems within their household. Thus, their increased awareness of problems translates to a decline in the measurable effects of the program. Second, older children may be treated differently, affecting the caregivers' perspective of the relationship between parent and child. Youth are one or two years older at endline than at baseline. As children get older, we heard from counselors that caregivers begin giving them more independence. That may explain why both treatment and control caregivers decline on the caregiver presence index from baseline to endline.

EX-POST MINIMUM DETECTABLE EFFECT

Table 10 reports ex-post MDE calculations for caregiver indicators. Column 1 shows the baseline mean for control caregivers in the matched TOT sample across each indicator. Column 2 reports the minimum detectable effect that the random effects model would reveal using the matched TOT sample. Column 3 shows the baseline mean for control caregivers in the TOT sample across each indicator. Column 4 reports the minimum detectable effect that the random effects model would reveal using the TOT sample.

We want to know if either sample would enable us to capture statistically significant results given the coefficients that we see in the main results. The MDE on parental authority is smaller than the coefficient on parental authority, pointing to the likelihood that the program does not impact caregivers' parental authority. For the remaining caregiver outcomes, the MDEs are larger than the coefficients on the results. Six of the eight remaining coefficients have an absolute value of less than 0.10 which would mean that even if evaluation was powered to capture such results, they would be quite minimal. Caregiver presence index and informed parent index have slightly larger, though negative, coefficients that the evaluation would not be able to capture. Overall, though not smaller in magnitude than many of the resulting coefficients, the MDEs are quite small compared to the means for each outcome, meaning we would be able to capture effects that would express a meaningful relationship between the program and family outcomes.

Table 10: Ex-post MDE for caregiver indicators

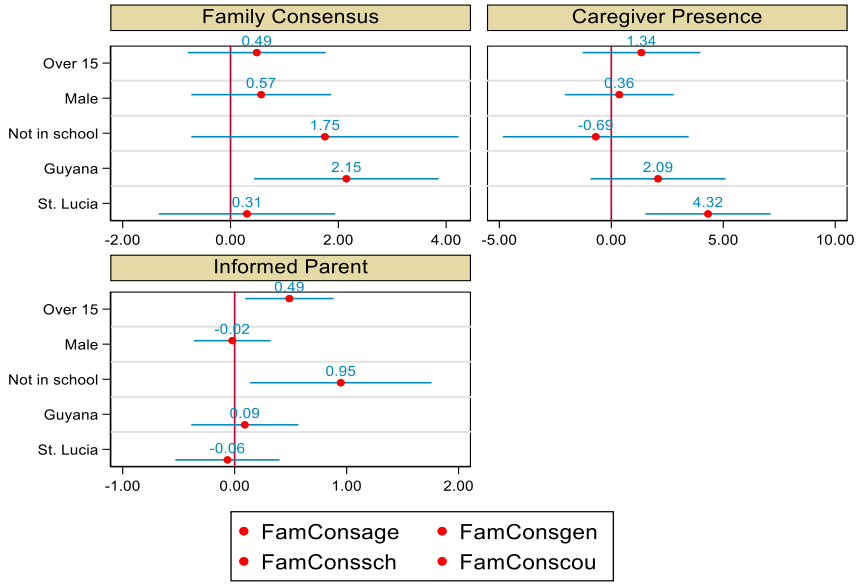
	Matched TOT		TOT	
	Control Mean at Baseline (n=293)	MDE	Control Mean at Baseline (n=470)	MDE
Extended Family Cohesion Index	8.8	0.39	9	0.33
Family Cohesion Index	0.017	0.11	0.04	0.10
Family Consensus Index	5.5	0.37	5.7	0.32
Parenting Locus of Control Index	13.5	0.25	13.5	0.21
Improvement Self-Evaluation	8.3	0.21	8.3	0.16
Caregiver Presence Index	11	0.86	11	0.70
Informed Parent Index	3.9	0.13	4	0.11
Parental Authority Index	21.6	0.41	21.7	0.35
Youth Behavior and Relationships Index	0.7	0.15	0.7	0.13

HETEROGENEOUS TREATMENT EFFECTS

The impact of the *Family Matters* program on the caregiver outcomes of interest may vary by group. Programmatically, such varied impacts may inform improved program targeting, and more generally, varied impacts help us interpret the overall results. As above, to understand variation in how different groups respond to *Family Matters*, we examine potential differences in program effects by 1) age group, 2) gender, 3) school enrollment status, and 4) country.

Figure 19 presents differential program effects on family consensus, caregiver presence and informed parent indices by varied groups, decomposed by these factors. There are some differential effects based on age, education and geographic location. We find a greater treatment effect in Guyana on building family consensus than in St. Kitts and Nevis, which is statistically significant at the 95 percent level. We also find a greater treatment effect on caregivers in St. Lucia in terms of caregiver presence as compared with St. Kitts and Nevis, which is statistically significant at the 99 percent level. Caregivers whose youth is over 15 years old and caregivers whose youth are out of school, on average, experience a greater treatment effect on the informed parent index than those whose youth is under 15 years old and those whose youth are in school. These differences are statistically significant at the 95 percent level.

Figure 19: Heterogeneous treatment effects on Family Consensus Index, Caregiver Presence Index, and Informed Parent Index



ASSESSMENT OF FAMILY DYNAMICS

While the indices are helpful from an evaluation perspective, they do not convey much information about the status of family dynamics nor provide a detailed picture of how these factors may or may not be changing overtime. To better illustrate the findings, as above with the YSET, we present the family cohesion index in greater detail. Many of the changes that could presumably be generated by the intervention occur at the level of family functioning and relationships, but these changes may not necessarily be reflected in explicit changes in youth attitudes and behaviors. Instead, changes in family communication and interpersonal relationships are likely to be reflected in Family cohesion index, which includes questions about the characteristics of family relationships, the family's ability to navigate challenges, and other aspects of family cohesion. To assess these changes, we analyze differences between responses to questions that comprise the Family cohesion index.

Figure 20 and

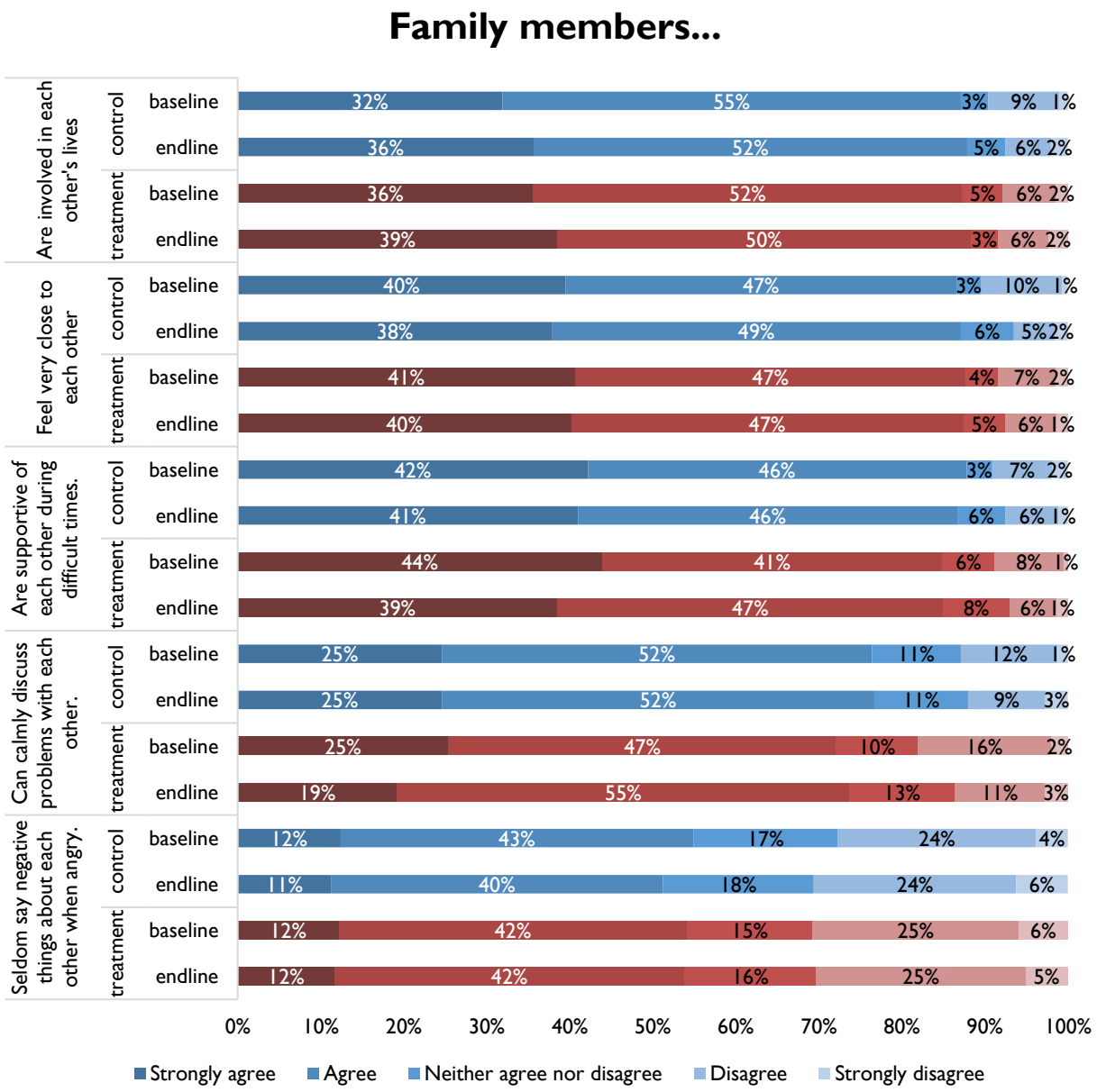
Figure 21 depict caregiver responses to Likert scale questions that comprise the Family cohesion index. Figure 20 shows to what extent caregivers agree that family members are involved in each other's lives, feel close to each other, are supportive of each other during difficult times, can calmly discuss their problems with each other, and rarely say negative things about each other. Figure 21 illustrates caregiver satisfaction with the quality of communication between family members, the family's ability to resolve conflicts, and the amount of time spent together as a family.

In comparing baseline and endline responses, caregivers in both the control and treatment groups were more likely to report that they strongly agreed with the statement "family members are involved in each other's lives" at endline (from 32 percent of control respondents and 36 percent of treatment respondents to 36 percent and 39 percent, respectively). However, in both groups, the overall levels of agreement and

disagreement did not significantly change. In the control group, the percent of caregivers who either agreed or strongly agreed with this statement only shifted from 87 percent to 88 percent, and in the treatment group, the shift was similarly small, increasing only by one percentage point from 88 percent to 89 percent. Interestingly, when asked to what extent they agreed with the statement "family members are supportive of each other during difficult times," the responses of caregivers in the control group remained largely unchanged, while the responses of those in the treatment group reflected less intensity of agreement at endline compared to baseline. At baseline, a relative majority of caregivers (44 percent) reported strong agreement with this statement, but at endline only 39 percent reported the same. Instead, it was more common for respondents to report that they just "agreed" at endline (47 percent), which reflects an increase of six percentage points from baseline.

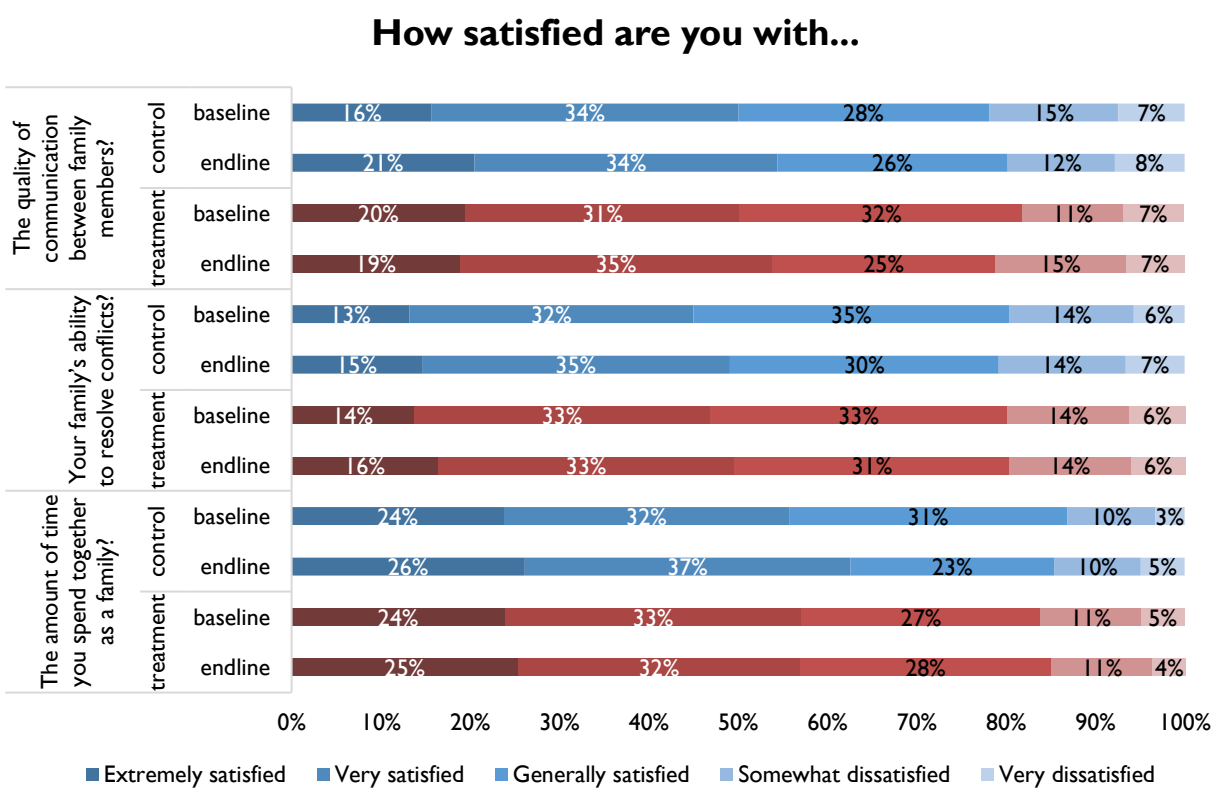
A similar shift occurred when caregivers were asked about the extent to which they agree with the statement "family members can calmly discuss their problems with each other." Again, we see little to no changes in the control group (other than three percent of respondents changing their answer from "disagree" to "strongly disagree"), but we see a marked shift in the treatment group, with less instances of respondents indicating they "strongly agree" and more reporting that they just "agree." As shown in Figure 20, at baseline, 47 percent of respondents in the treatment group said they agreed with the statement and 25 percent said they strongly agreed. By endline, the number of respondents who agreed increased to 55 percent, and the number who strongly agreed dropped to 19 percent. However, despite this dip in enthusiasm, the treatment group also saw less overall disagreement with the statement at endline. At baseline, 18 percent reported they either strongly disagreed or disagreed with the statement, but at endline, this number decreased to 14 percent. Instead, there was an increase in the number of respondents in the treatment group who neither agreed nor disagreed with the statement from baseline to endline (from 10 percent to 13 percent).

Figure 20: Characteristics of relationships among family members



When asked about their satisfaction with the level of communication between family members, in both control and treatment groups, caregivers were more likely to report they were either "extremely satisfied" or "very satisfied" at endline compared to baseline (from 50 percent of control respondents and 51 percent of treatment respondents to 55 percent and 54 percent, respectively). However, the two groups were dissimilar in terms of their responses at the other end of the scale. As shown in Figure 21, the number of respondents who were dissatisfied decreased in the control group from baseline to endline (22 percent to 18 percent) but increased in the treatment group over the same period (20 percent to 22 percent).

Figure 21: Caregiver satisfaction with elements of family cohesion



From the caregivers' perspective, it does not appear that family dynamics have changed significantly since baseline. However, among caregivers in the treatment group, there is a more noticeable trend of responses shifting from a more positive tone to a more negative one (i.e., "strongly agree" to. "agree," "somewhat dissatisfied to "very dissatisfied"). During qualitative fieldwork, counselors mentioned that initially many families are reluctant to see or admit they have problems with their children. For example, caregivers originally did not think it was a problem to not know where their children are. As they move throughout the program, however, they gradually start to see these as problems. As a result of this change in perspective, they may be more aware of their challenges and more likely to report them at endline, even if improvements have been made, which can explain the downward shift in agreement and satisfaction with measures of family cohesion seen in the treatment group.

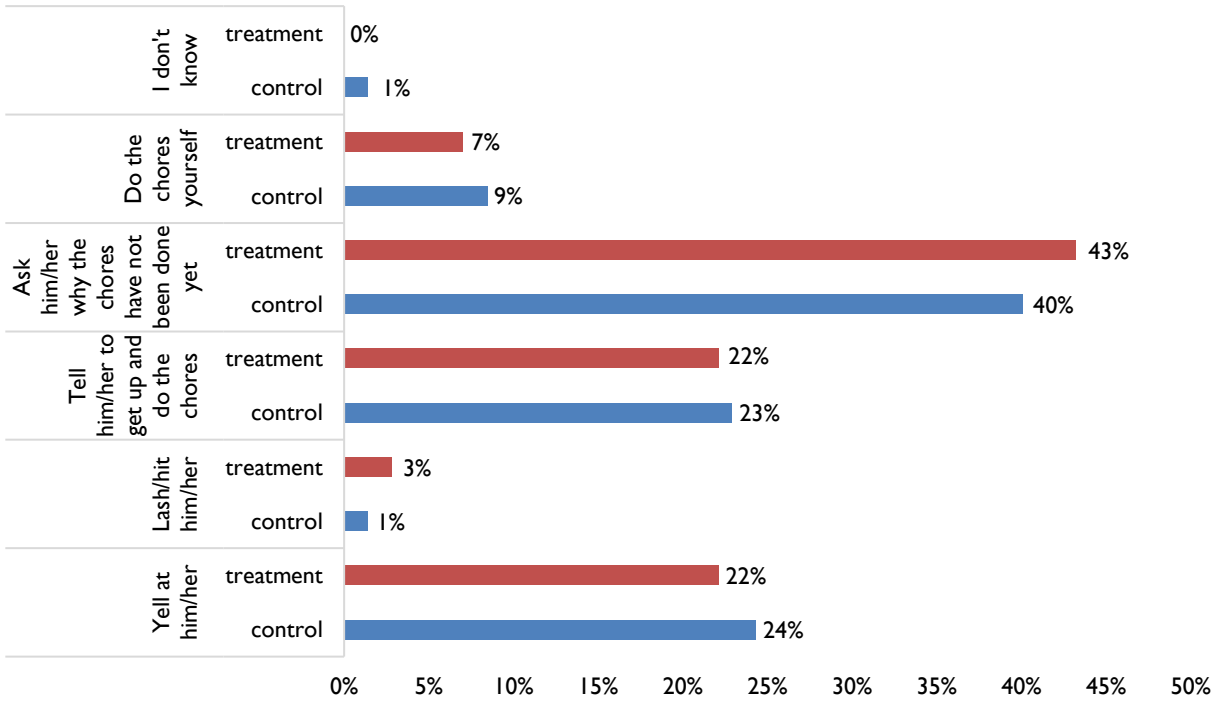
CAREGIVER RESPONSES TO YOUTH BEHAVIOR

A key caregiver behavior that counselors sought to change was caregivers' responses to challenging family situations, such as their child being rude or misbehaving. The goal is for caregivers to adjust their response away from physical or verbal violence and toward dialogue with the youth. To gauge caregiver response and behavior, we posed a series of scenarios to treatment and control caregivers. Scenarios were only included in the endline survey, so we cannot compare caregiver changes over time. The idea for these scenarios emerged from our research trips to the three countries, in which we conducted focus group discussions with youth and separately with caregivers, and interviewed family counselors and other key stakeholders. After hearing about the types of situations that caregivers found challenging and learning that counselors were working with them to improve their response, we designed these two scenarios to

try to detect differences in response between treatment and control caregivers. These conversations illuminated the fact that the baseline caregiver survey did not capture behaviors that counselors were working with families to improve, so the additional questions and results supplement the main findings.

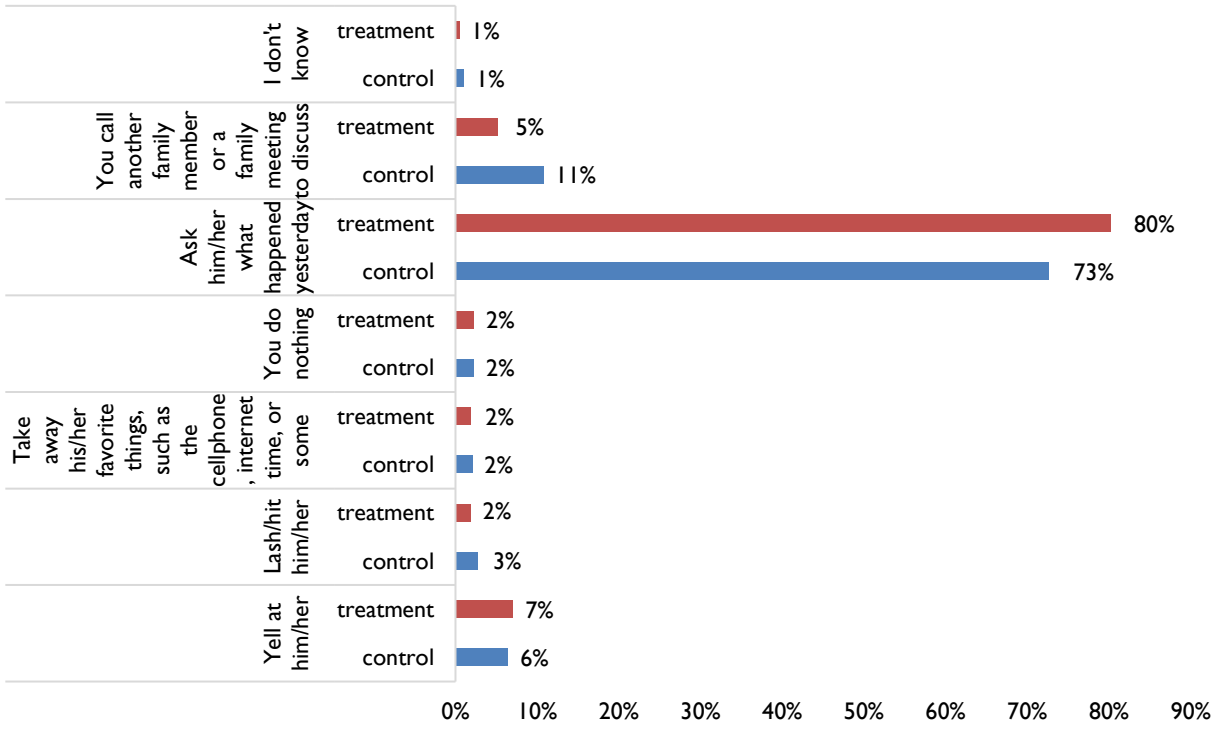
Figure 22, Figure 23, and Figure 24 report caregiver responses to each scenario disaggregated by treatment and control. In the first scenario, caregivers are asked what they would do if they returned home after asking their child to do their chores, only to see that the chores have yet to be done. The relative majority of both treatment and control caregivers would ask about the chores. Treatment caregivers are more likely to ask youth about the chores and less likely to yell at the youth than control caregivers. However, three percent of treatment caregivers respond that they would lash or hit their child as opposed to one percent of control caregivers.

Figure 22: Caregiver response to scenario where youth did not do chores



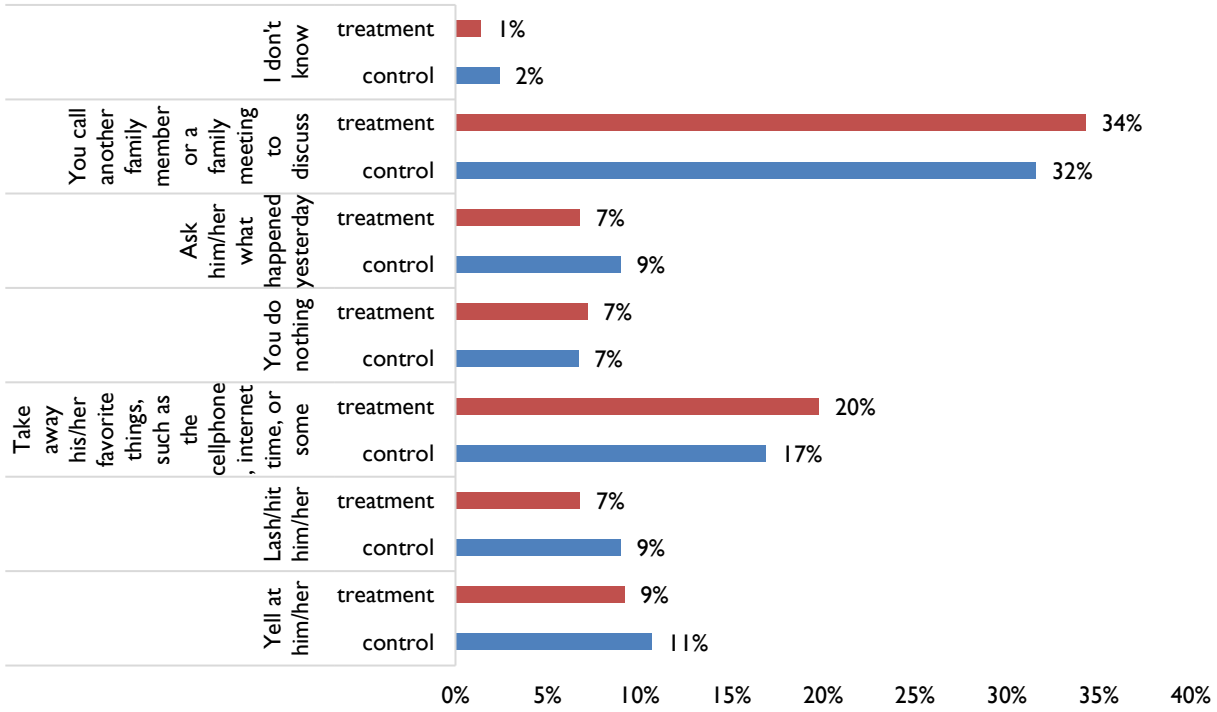
The next scenario asks what the caregiver would do after hearing from a neighbor that they saw the youth skipping school or work and smoking marijuana. The majority of both treatment and control caregivers would ask the youth what happened, but seven percent more treatment caregivers than control caregivers would ask the youth what happened. One more percent of treatment caregivers would lash or hit the youth than control caregivers, but the percentage of each who said they would lash or hit the child is low (seven and six percent, respectively).

Figure 23: Caregiver response to scenario where neighbor reports that youth skipped school or work to smoke marijuana



Finally, we ask caregivers what step they would take next (after their initial response to their youth regarding skipping school or work and using marijuana). The relative majority of both treatment and control caregivers would call a family meeting to discuss the incident, but two percent more treatment caregivers respond that they would take this approach than control caregivers. Fewer treatment caregivers than control caregivers would either lash or hit the child or yell at the child than control caregivers.

Figure 24: Caregiver response about what he or she does next after the prior scenario



While these results do not definitively point to the program impacting caregiver behavior, they are encouraging. Even though the differences are often slight, more treatment caregivers than control caregivers seek to communicate with their child regarding behavior problems and fewer respond violently. Such results could mean that the program provided caregivers with sound rationale for using dialogue as a means to remedy disputes with their children and provided them with the tools to do so.

FAMILY RISK DETERMINANT ANALYSIS

We explore how family structures relate to risky behaviors in youth to better understand program targeting and take up. This is not part of the evaluation per se, but rather an exploratory analysis to better understand the nature of risk in the targeted countries. Specifically, we consider four family types: nuclear families where both a mother and father are present; single father households where only a father is present, single mother households where only a mother is present; and households where neither a mother nor father are present (most children with no parents do have some adult caregiver in the household, most often a grandmother). At baseline, there were 276 nuclear families, 85 single father households, 520 single mother households and 121 households with neither a mother nor father.

To examine relationships between family type and youth risk, we run a regression of risk factor on *nuclear family* with dummy variables for *single father*, *single mother* and *no parents* at baseline. Table 11 shows relationships between family types and youth risk factors. Column 1 reports the coefficient and standard error of risk factors on *single father* compared to *nuclear family*. Column 2 reports the coefficient and standard error of risk factors on *single mother* compared to *nuclear family*. Column 3 reports the coefficient and standard error of risk factors on *no parents* compared to *nuclear family*. Column 4 reports the R-squared value, or the proportion of the variance of the risk factor explained by family types.

Family structure is generally not a statistically significant predictor of overall risk or specific risk factors. While few differences are statistically significant, the direction of the coefficients do reveal some trends. Children in households where both parents are present appear least risky. Children from single mother households are more likely than children from nuclear households to score at risk on six of the 11 measures of risk than children in non-single mother households. Children from single father households are more likely to score at risk on seven of the 11 measures of risk than children in non-single father households. Two of these differences, peer delinquency and family anti-social influence, are statistically significant at the 90 percent and 95 percent level, respectively. Children from households without a mother or father appear riskiest and are more likely to score at risk on nearly all measures of risk than children in dual parent households.

Table 11: Family risk determinant

	Single Father (N = 85)	**Single Mother (N = 520)**	**No Parents (N = 121)**	**R-squared**
Four or more risk factors	-0.059* (0.03)	0.011 -0.02	0.014 (0.03)	0.006
Mean number of risk factors	0.182 (0.19)	0.095 (0.11)	0.309* (0.17)	0.004
B: Weak parental monitoring	-0.043 (0.07)	-0.051 (0.04)	0.013 (0.06)	0.003
C: Critical life events	0.049 (0.06)	0.047 (0.04)	0.061 (0.06)	0.002
DE: Impulsive risk taking	-0.043 (0.06)	0.000 (0.03)	0.069 (0.05)	0.004
EG: Risky behaviors	0.056 (0.06)	-0.015 (0.04)	0.046 (0.06)	0.003
F: Guilt neutralization	-0.142*** (0.07)	-0.042 (0.03)	-0.056 (0.05)	0.009
G: Negative peer influence	0.000 (0.07)	-0.009 (0.04)	-0.003 (0.06)	0.000
H: Peer delinquency	0.123* (0.07)	0.009 (0.04)	0.029 (0.06)	0.004
IJ: Self-reported delinquency	0.047 (0.06)	0.047 (0.04)	0.055 (0.05)	0.002
T: Family antisocial influence	0.134** (0.07)	0.108*** (0.04)	0.094 (0.06)	0.010

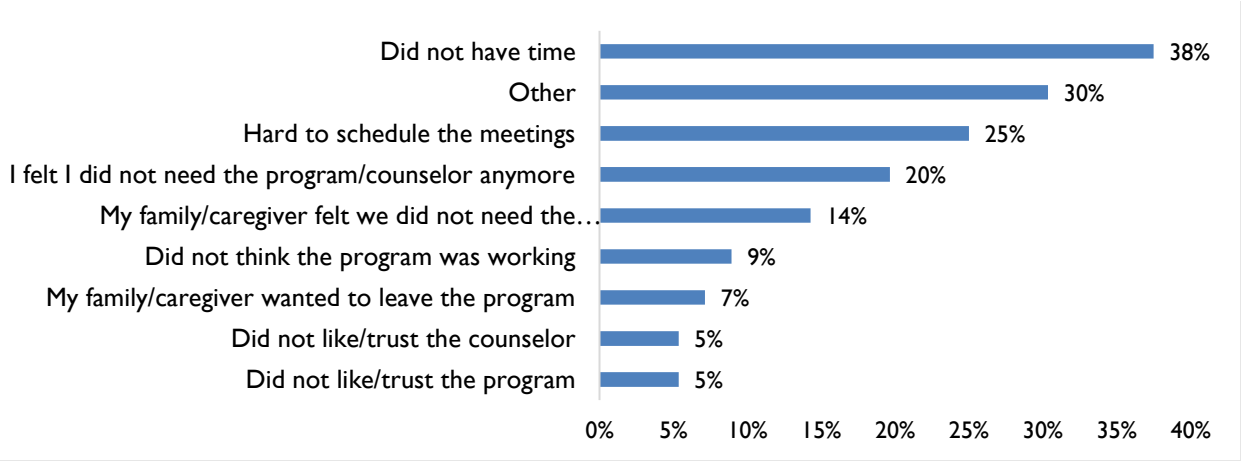
Note: *** p<0.01, ** p<0.05, * p<0.1

YOUTH ASSESSMENT OF THE PROGRAM

Treatment youth were asked a series of questions about their opinion of their family counselor and the *Family Matters* program in general. Despite the limited impact identified above, treatment youth generally offer a very positive view of the counselors and the program.

With youth who started but did not complete the program, we asked them why, to which the most common answer was "did not have time" (38 percent). Only nine percent said the program was not working. The main reasons for not completing the program appear to be logistical and due to scheduling issues rather than program dissatisfaction.

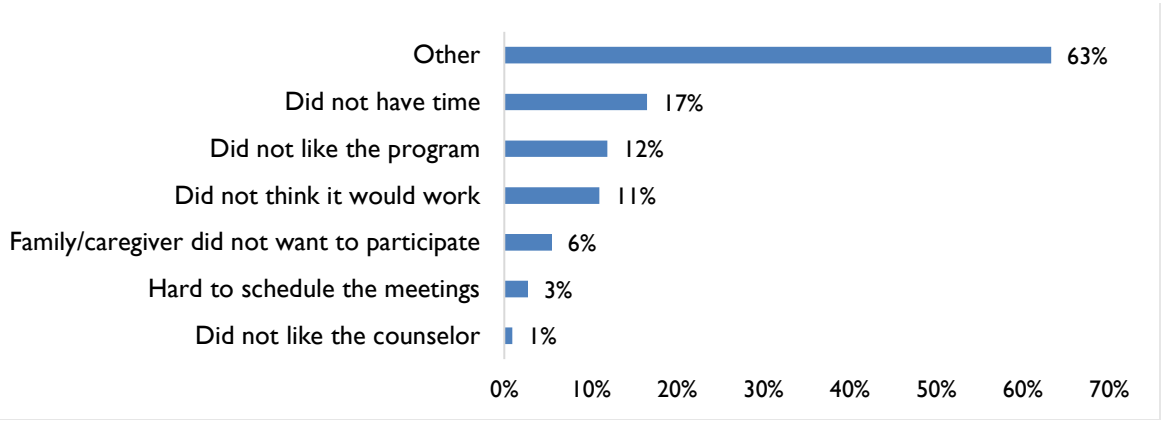
Figure 25: Treatment youth responses to the question "why did you not complete the program?" (n=56)



With youth who were invited but refused to participate in the program, we also asked why they refused. As seen in

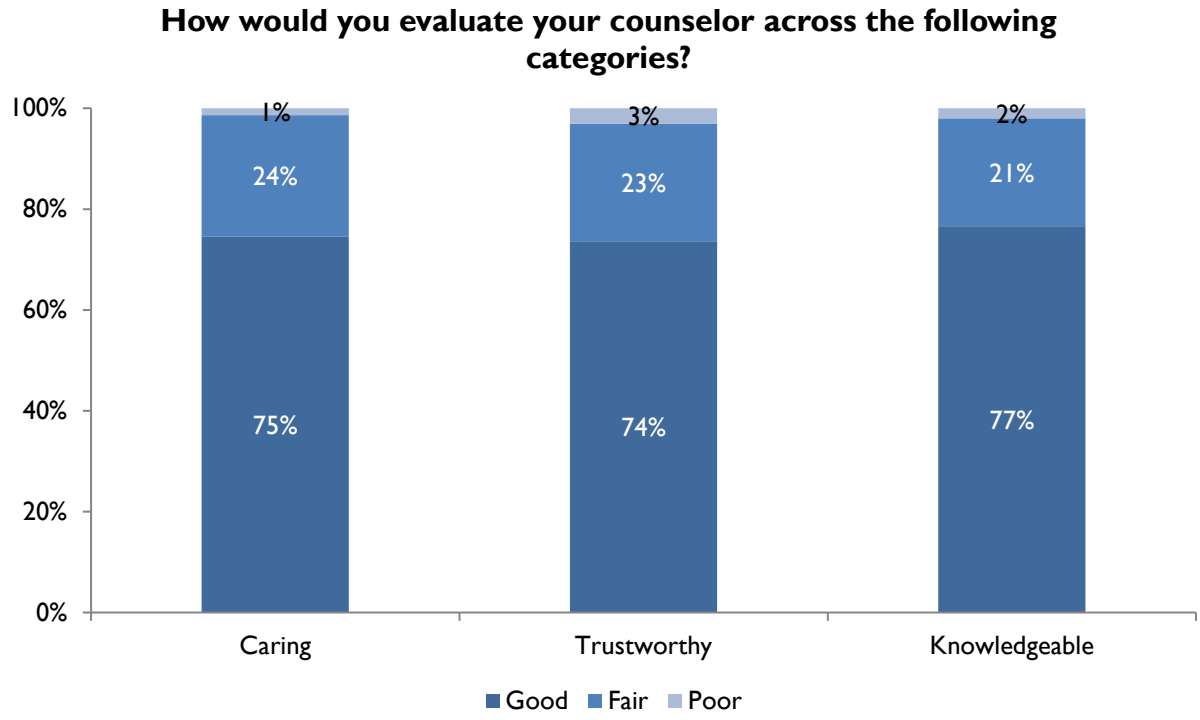
Figure 26, a large majority of respondents selected "other" rather than the provided options, including lack of time or disinterest in the program. The majority of those who coded other reported that they did not know about the program. It is unclear if they cannot recall hearing about the program, if somebody else in the family made the decision to decline participation, or if greater outreach could have been warranted. Caregivers were the ones approached about participating in the program, and they or may not have discussed this with the youth, which may explain why a large share of youth coded as "declined participation" report not knowing about the program.

Figure 26: Treatment youth responses to the question "why did you not participate in the program?" (n=109)



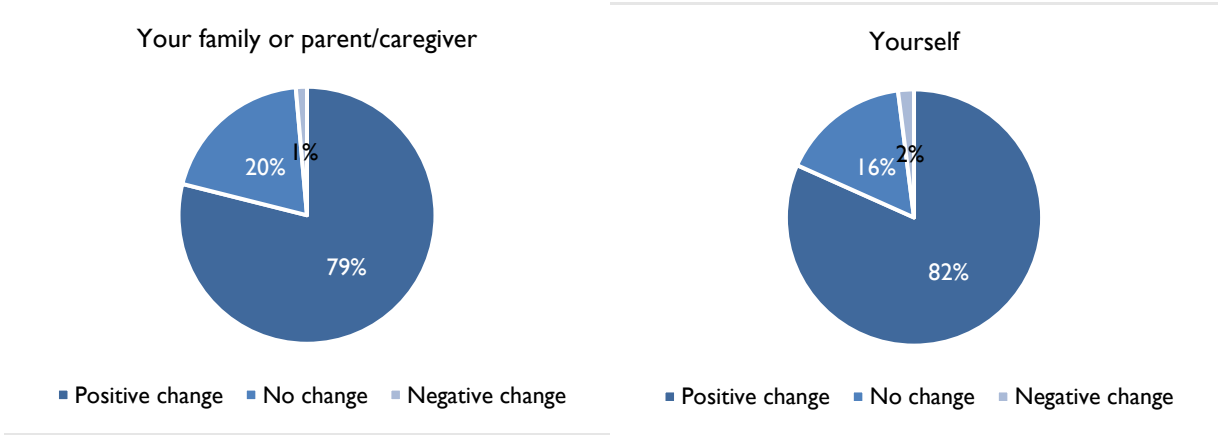
Treatment youth who participated in the program were asked to rate their counselors along the caring, trustworthy and knowledgeable scales. About three-fourths of the respondents gave their counselors the highest rating of "good" on each of those categories.

Figure 27: Treatment youth responses to evaluating their counselors (n=295)



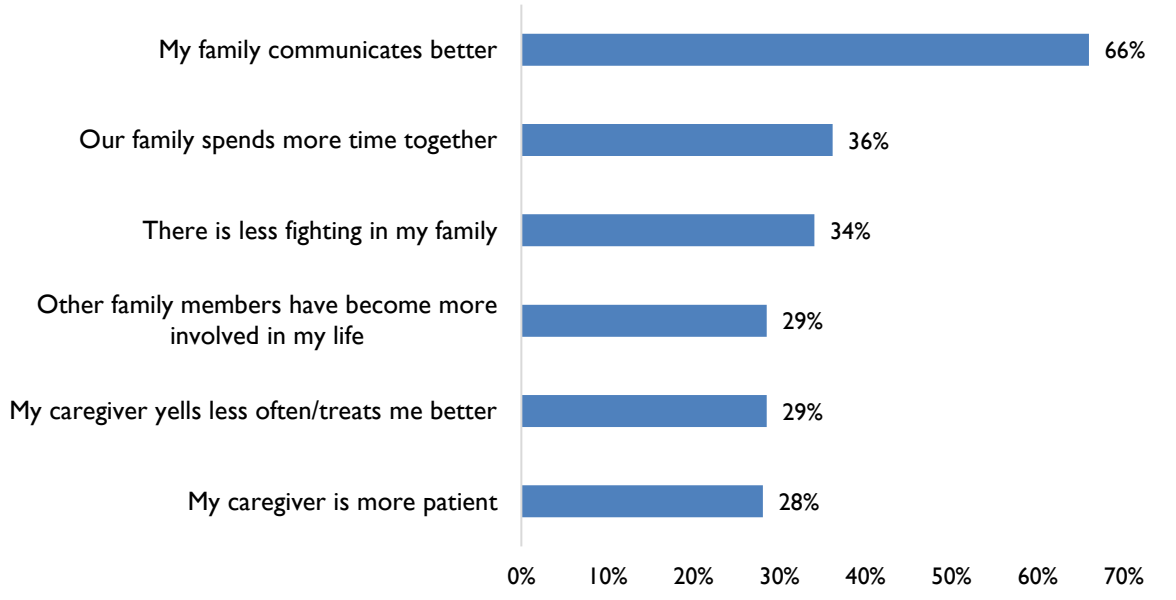
When we asked treatment youth whether counseling has caused a change in their family or caregiver, 79 percent of youth said counseling had a positive change on their families and caregivers, while 20 percent reported no change. When asked the same question about a change in themselves, a greater number (82 percent) reported positive changes.

Figure 28: Do you think that the counseling has had a positive change, a negative change, or no change on... (n=295)



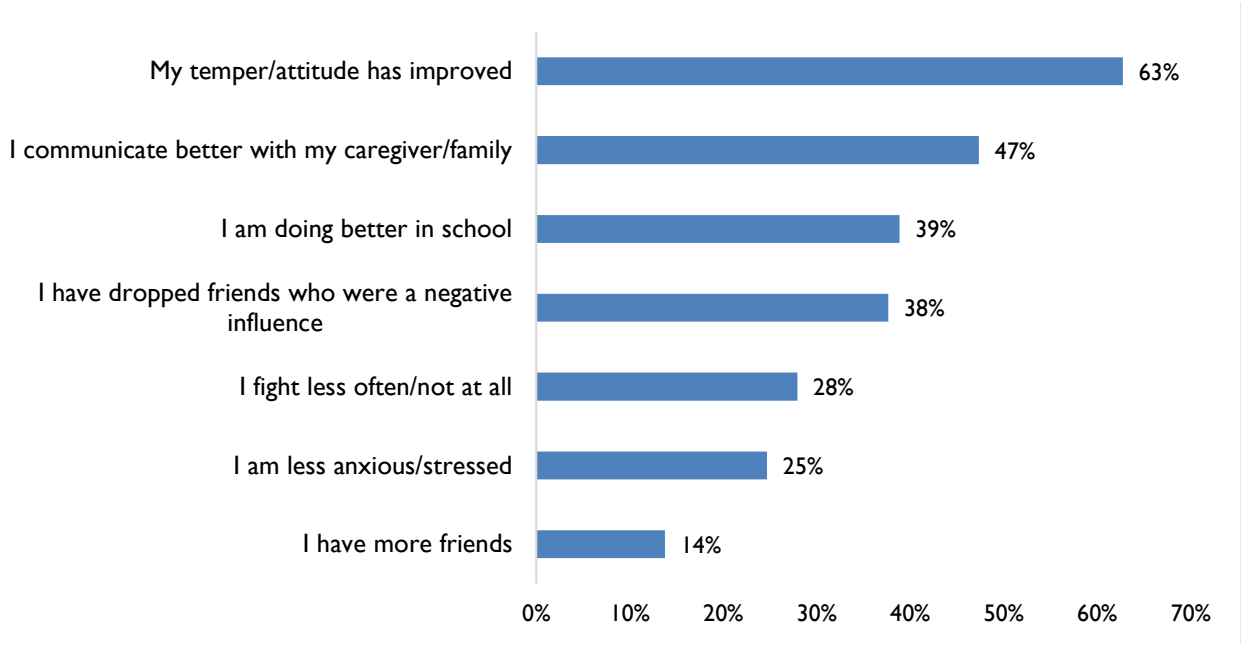
The majority (66 percent) of those who said there have been positive changes in their families/caregivers reported that the family communicates better now. About 30 percent of youth reported each of the other positive changes, including the family spending more time together, less fighting in the family, other family members becoming more involved in the youth's life, caregivers yelling less and treating the youth better, and caregivers being more patient.

Figure 29: Positive changes youth have seen in their families/caregivers (n=235)



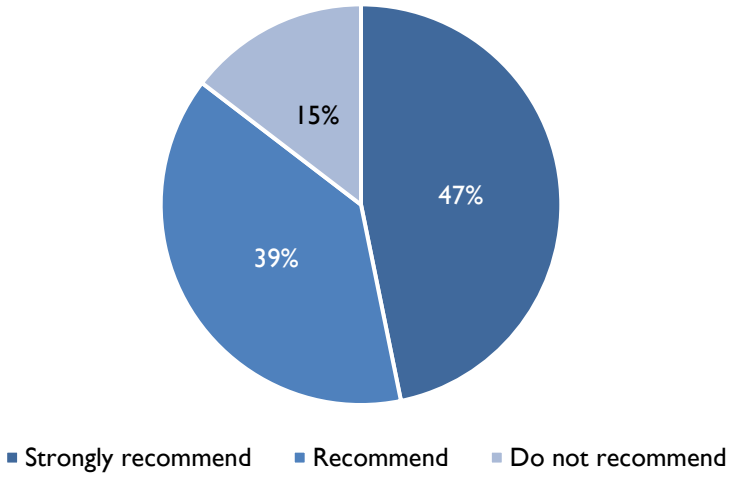
63 percent of youth who said they have seen positive changes within themselves reported an improved temper/attitude, and 47 percent reported better communication with the caregiver/family.

Figure 30: Positive changes youth have seen in themselves (n=247)



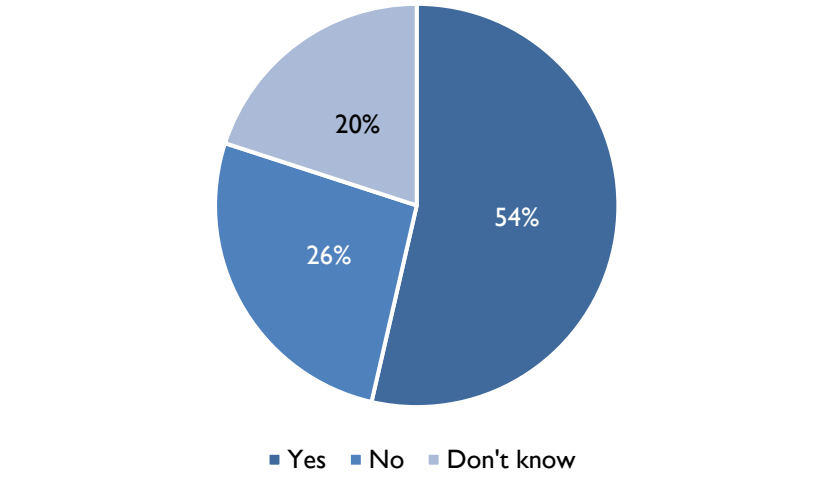
Most youth (86 percent) who participated in the program would recommend it to their friends, and 47 percent would strongly recommend it.

Figure 31: Whether youth would recommend the program to their friends (n=295)



To those who reported having friends who are also enrolled in the *Family Matters* program, we asked whether they thought these friends were less likely to get in trouble now. More than half (54 percent) said yes, 26 percent said no, and 20 percent said they did not know.

Figure 32: Whether their friends who are also enrolled in the program are now less likely to get in trouble at school or at home, because of the program (n=125)



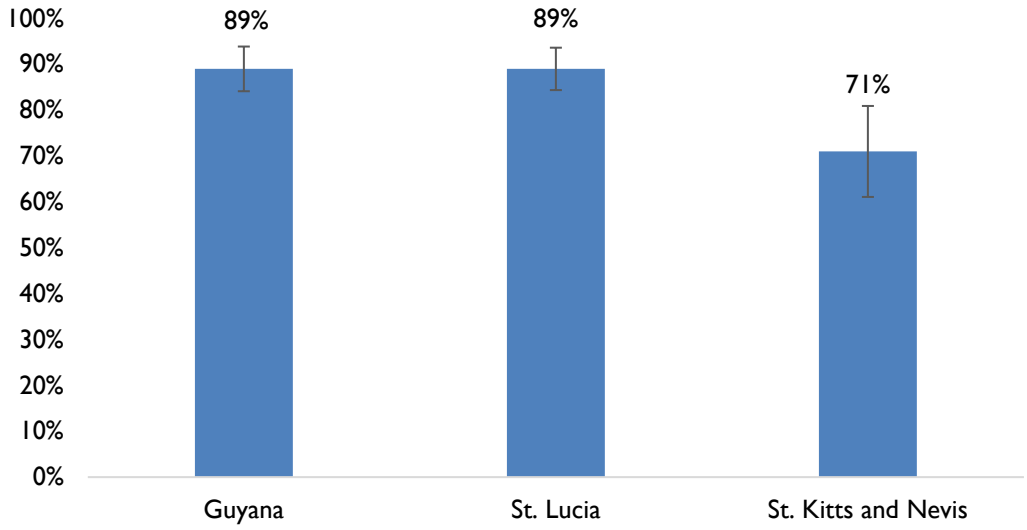
Overall, the youth assessment of the program is quite positive. Most youth saw positive changes in themselves and their parents/caregivers, rated their counselors highly, and would recommend the program to others.

DIFFERENCE IN PROGRAM RECOMMENDATION BY GENDER, AGE GROUP AND COUNTRY

We also checked whether there are differences across gender, age group or country. There are not significant differences between males and females, nor between older and younger respondents, but when comparing across countries, we find that respondents in St. Kitts and Nevis were less likely to recommend the program. Figure 33 shows the share of youth respondents who said they would recommend the *Family Matters* program to their friends by country (the answers "recommend" and "strongly recommend" were aggregated for this chart). While 89 percent of youth in Guyana and St. Lucia said they would recommend it to their friends, only 71 percent in St. Kitts and Nevis said so.

Not recommending the program does not seem to be a result of increasing risk levels from baseline to endline (according to the YSET), as 19 percent of SKN youth whose risk increased said they would not recommend the program, compared with 30 percent of those whose risk declined. We checked the other program assessment questions to figure out where this country difference might come from, but we did not find significant differences between SKN respondents and respondents from other countries in terms of counselors' ratings or the share of those who saw changes in themselves or their families/caregivers. It is unclear exactly why respondents in St. Kitts and Nevis are less likely to recommend the program. Our impressions during the midline qualitative research trips did not indicate that youth in St. Kitts and Nevis were less satisfied with the program than in the other two countries.

Figure 33: Share of youth respondents who would recommend or strongly recommend the program to their friends, by country (n=295)



CAREGIVER ASSESSMENT OF THE PROGRAM

We also asked treatment caregivers direct questions regarding their experience with *Family Matters* counselors and how their family dynamics may have changed as a result of the program. Similar to youth and despite the impact evaluation findings, caregivers generally evaluated the program well and perceived it to be effective.

Figure 34 displays explanations for why caregivers who were offered to participate in the program failed to complete the program. Most caregivers selected an answer outside of the answer choices provided, reflecting the wide variation in how families make decisions. Unfortunately, we did not ask participants to clarify the "other" reason as was done with the YSET; however, it seems reasonable to assume that it was the same issue of unawareness of the program. The next most common answers concerned their availability. 29 percent of caregivers dropped out of the program because it was difficult to schedule meetings and 27 percent dropped out because they did not have time to participate in the program.

Figure 34: Treatment caregiver's responses to the question "why did you not complete the program?" (n=66)

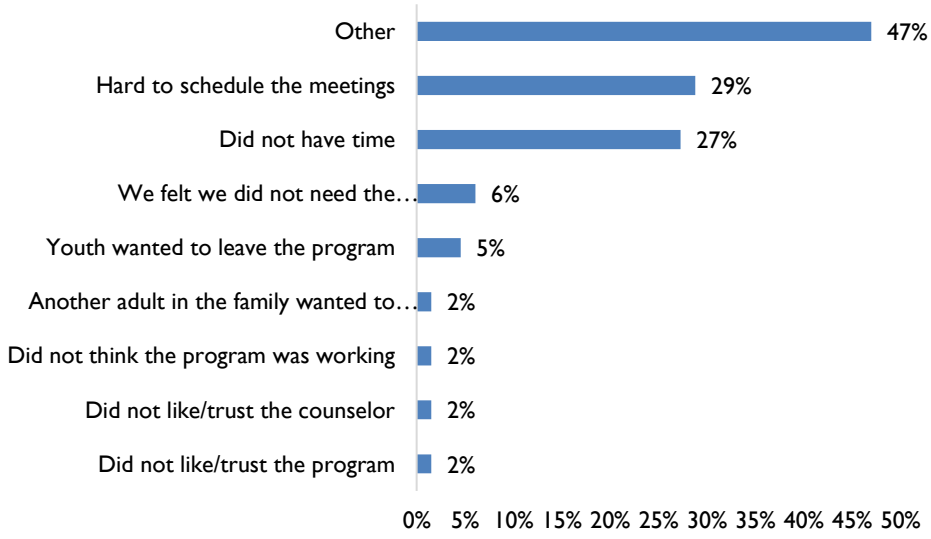


Figure 35 shows explanations for why treatment caregivers failed to complete the program or declined to participate in the program from the beginning. 18 percent of caregivers declined to participate in the program because they did not have time to participate in the program. Few caregivers failed to participate in the program because they did not like or trust the counselor or the program, which tells us that caregivers, at best, have a positive view of *Family Matters*, and at worst, a neutral view of the program.

Figure 35: Treatment caregiver's responses to the question "why did you not participate in the program?" (n=73)

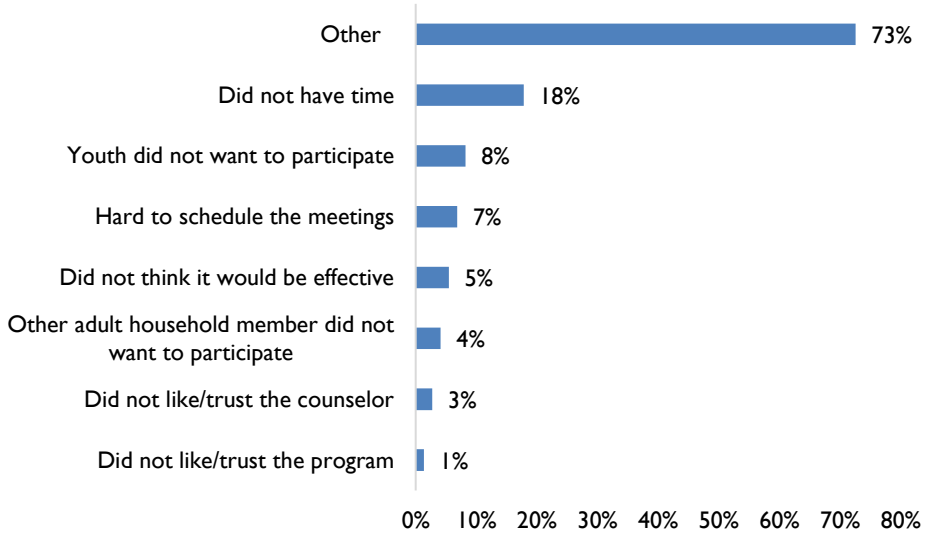


Figure 36 shows caregivers' evaluation of *Family Matters* counselors. The vast majority of caregivers (89 to 92 percent) rate the counselors as good across the scales of knowledgeable, caring, accessible and trustworthy. One percent or fewer caregivers rate the counselors as poor across the scales of counselor characteristics.

Figure 36: Treatment caregivers' evaluating their counselors across categories including knowledgeable, caring, accessible and trustworthy (n=283)

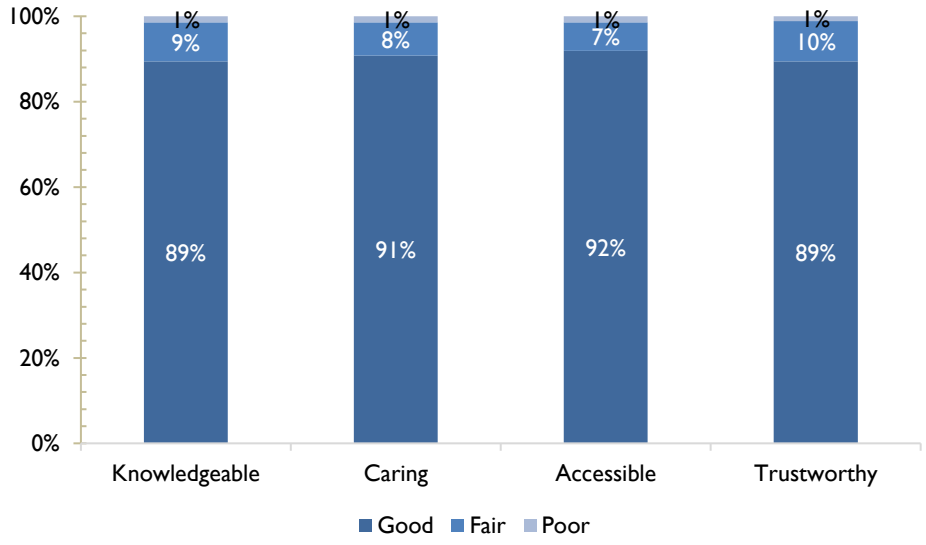


Figure 37, Figure 38, Figure 39, and Figure 40 reveal caregivers' perspectives on positive changes the program has had on their family and child. The majority of caregivers believe the program had a very positive or positive change on both their child and family. 10 percent of caregivers saw no change in their child or family, and one percent or fewer of caregivers saw a negative change in their child or family. While the majority of all groups report that the program had a positive impact on their child or families, female caregivers are slightly more likely to report positive changes than male caregivers. Caregivers from St. Kitts and Nevis and Guyana are also slightly more likely than caregivers from St. Lucia to report positive changes. The largest differences in opinions about the positive impact of the program exist among different family types. Single father households are less likely than other household types to report a positive impact of the program, with only 76 percent of caregivers from single father households reporting positive impacts on the child or family as a whole. While these statistics imply a generally positive perception of the program, the differences may point to a need to improve program targeting or tailoring for different needs of different types of caregivers or families.

Figure 37: Do you think that the counseling has had a very positive change, positive change, a negative change, or no change on... (n=283)

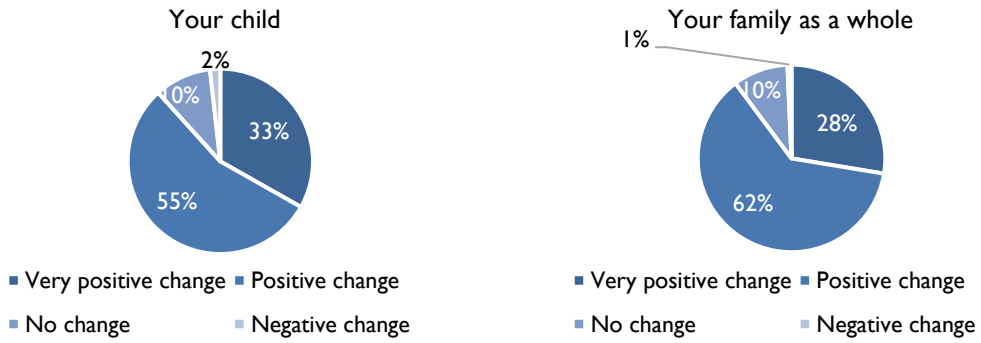
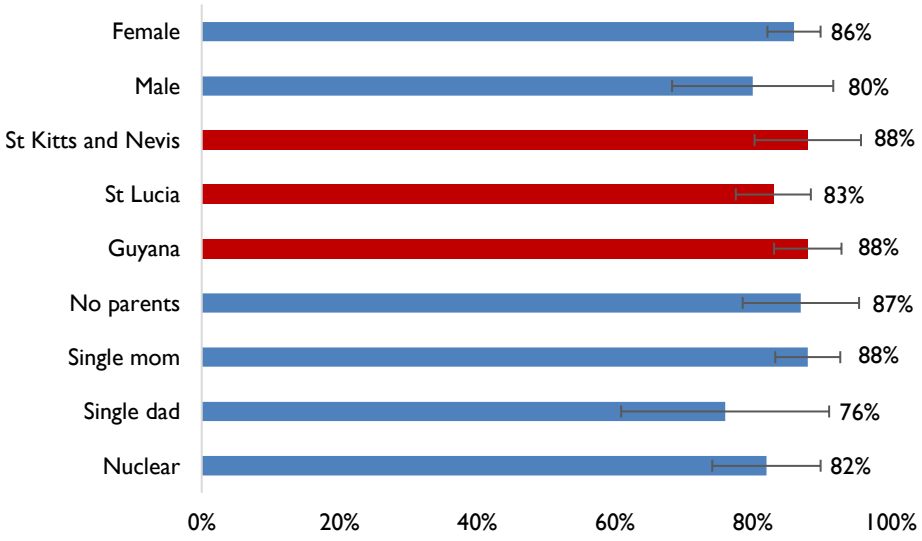
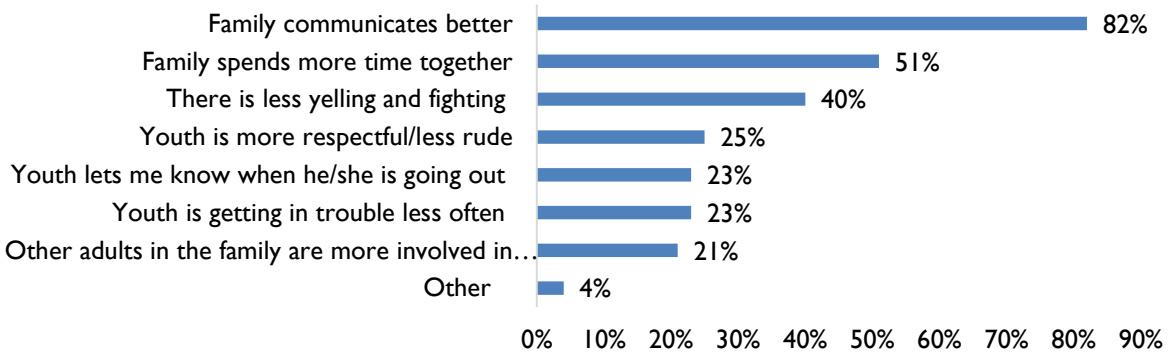


Figure 38: Percent of caregivers who say the program had a positive effect on their family and/or child disaggregated by gender, country and family type (n=283)



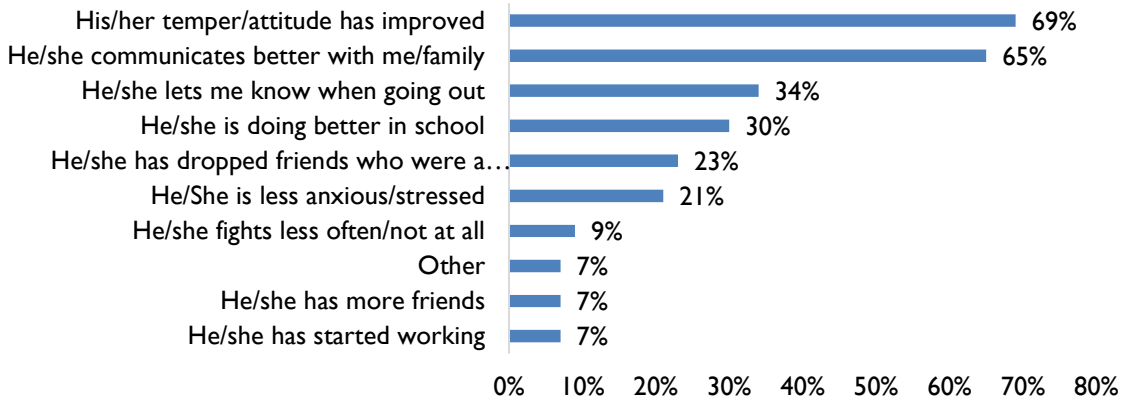
Of caregivers who responded that they see a positive change in their families, 82 percent report that their families communicate better; 51 percent report that the family spends more time together, and 40 percent report that there is less yelling and fighting. Far fewer (25 percent or below) caregivers report that the youth is more respectful, lets the caregiver know when he/she is going out, or get in trouble less. Additionally, few caregivers (21 percent) report that other adults in the family are more involved in the care for the youth.

Figure 39: Positive changes caregivers have seen on their families (n=254)



Of the caregivers who responded that they see a positive change in their child, 69 percent report that his/her temper or attitude has improved; 65 percent report that he/she communicates better. Fewer caregivers (23 percent or below) reported that the youth has dropped friends who were a negative influence, is less anxious or stressed, fights less often or not at all, has more friends or has started working. Positive improvements within the family and child appear to center around communication and general demeanor, as caregivers do not observe changes in risky behaviors.

Figure 40: Positive changes caregivers have seen on their child (n=250)



As noted above, this overwhelmingly positive perception of the program is not reflected in the program's impact on caregiver outcomes of interest. Across the caregiver outcomes, caregivers either do not improve or improve by such a small amount that the changes experienced by treatment caregivers are not significantly different from changes experienced by control caregivers.

Figure 41 and Figure 42 show caregivers' inclination to recommend the *Family Matters* program. 55 percent of caregivers strongly recommend the program, and 40 percent of caregivers recommend the program. Male caregivers and single father households are less likely to recommend the program than female caregivers and other household types.

Figure 41: Whether caregiver would recommend the program (n=283)

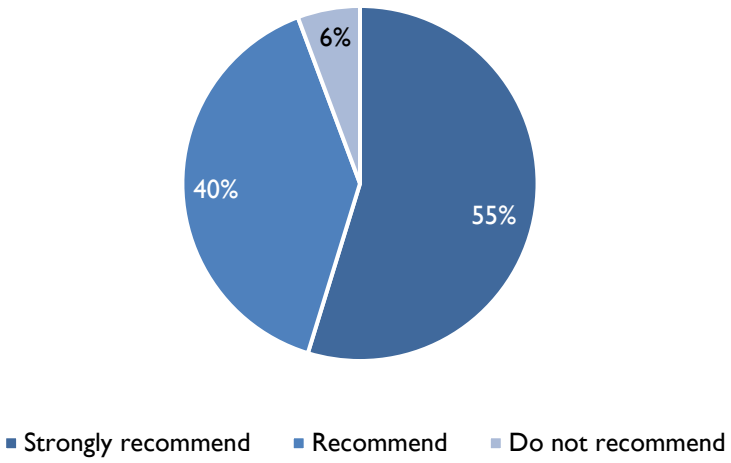
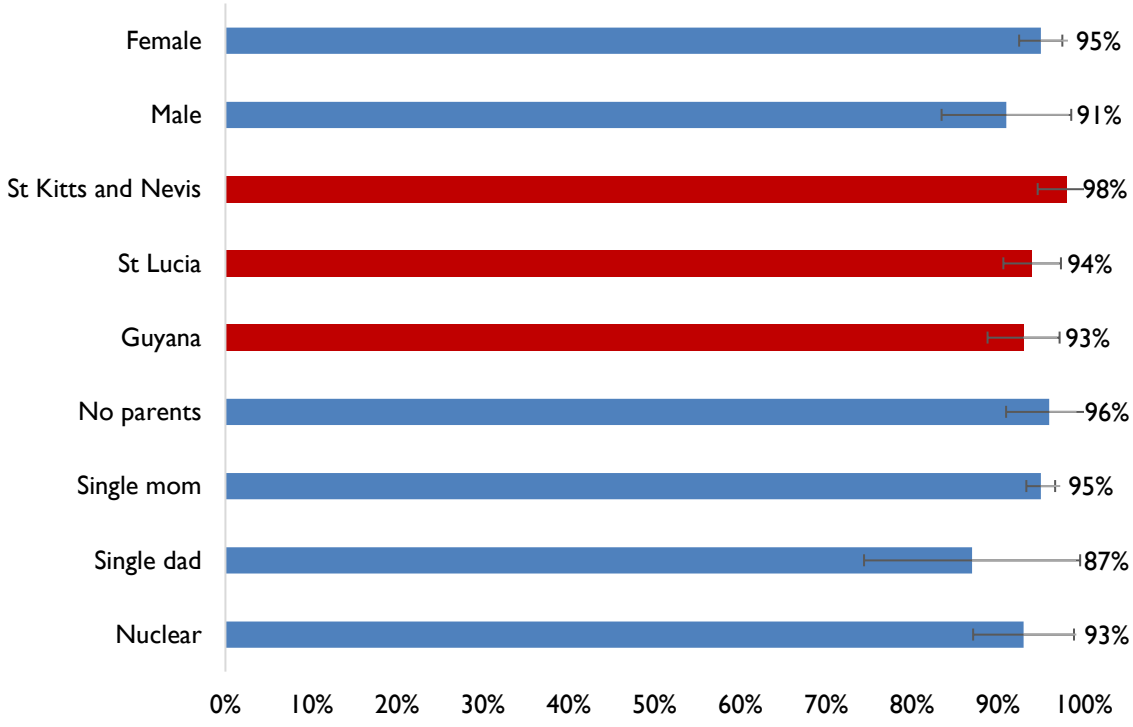


Figure 42: Percent of caregivers who would recommend the program disaggregated by gender, country and family type (n=283)



TARGETING ANALYSIS AND ATTRITION

YSET SURVEY ATTRITION

Out of the 566 youth who were randomized to treatment at baseline, 481 of them were re-surveyed at endline. Therefore, the treatment sample lost 15 percent of youth due to survey attrition. Out of the 567 youth that were randomized to control at baseline, 499 of them were re-surveyed at endline. This means that the control sample lost 12 percent of youth due to survey attrition. We examined if the youth lost to survey attrition are any different from those who remained in the survey sample.

Table 12 shows the results of a single regression of the binary variable *attrition_survey* (0=re-surveyed at endline, 1=dropped out of survey) on each of the youth characteristics and risk levels at baseline, with coefficients and standard errors reported in each cell.[15] Youth who dropped out of the survey were seven percent more likely to be older than 15 at baseline, 15 percent more likely to be out of school, and had higher scores in the overall YSET and in the EG – Risky behavior module.

[15] For a discussion of linear versus logistic regression when the dependent variable is a dichotomy, see Hellevik, 2009. We decided to use linear regression here because of the intuitive meaningfulness of the results.

Table 12: Results of a single regression of survey attrition on youth characteristics and risk level at baseline (n=1,131)

	Youth who dropped out of the survey
Male	-0.021 (-0.021)
Over 15 years old	0.069*** (-0.021)
Not in school	0.150*** (-0.033)
4 or more risk factors	-0.023 (-0.044)
Mean number of risk factors	0.018** (-0.007)
Youth has a mother/stepmother/female guardian	-0.049 (-0.08)
Youth has a father/stepfather/male guardian	-0.017 (-0.034)
B: Weak parental monitoring	0.034 (-0.021)
C: Critical life events	0.003 (-0.022)
DE: Impulsive risk taking	0.021 (-0.025)
EG: Risky behavior	0.037* (-0.022)
F: Guilt neutralization	0.011 (-0.027)
G: Negative peer influence	0.017 (-0.022)
H: Peer delinquency	0.02 (-0.021)
IJ: Self-reported delinquency	-0.004 (-0.024)
T: Family anti-social influence	0.019 (-0.021)

Note: Coefficient, standard error and level of statistical significance reported. * p<0.10, ** p<0.05, *** p<0.01

YSET PROGRAM ATTRITION

We also checked if there are statistically significant differences between: 1) youth who completed the program and youth who dropped out, and 2) youth who took up the program and youth who declined to participate. It is important to note that since *Family Matters* is a family-based intervention, it is not only the youth who decides whether or not to participate; a caregiver, representing the family, needs to provide consent/agree that the entire family will participate. In order to run this analysis, we ran a single regression of the binary variable *dropped_out* (0=completed treatment, 1=dropped out) on youth characteristics and risk levels at baseline, and we did the same for the binary variable *declined_participation* (0=took up treatment, 1=declined treatment). Table 13 shows the results of the single regressions on these two

outcome variables. Youth aged 15 and above at baseline were nine percent more likely to drop out of the program, and youth who reported being out of school at baseline were 24 percent more likely to drop out of the program. The finding that out-of-school youth were more likely to drop out of the program is concerning, considering that out-of-school youth are generally at a higher risk.

No other factors analyzed seemed to influence the likelihood of dropping out. Module IJ: self-reported delinquency is the only factor analyzed that predicts likelihood of declining to participate in the program. Youth who scored at risk on self-reported delinquency were 12 percent less likely to decline participation in the program. The difference is statistically significant at the one percent level.

We compare these attrition findings to those in the evaluation of the Los Angeles GRYD, from which the *Family Matters* program is derived. The GRYD evaluation found that older youth (aged 13 to 15) and those with more risk factors were less likely to graduate from the program (pp. 28-29 in Cahill et al, 2015). Similarly, we have found that older youth (ages 15 and above) were less likely to complete the *Family Matters* program. Risk factors as measured by the YSET do not appear to predict attrition in this study. However, being out of school, which is also a measure of risk, did increase the likelihood of dropping out of the *Family Matters* program. For GRYD, nearly all youth in the program were in school, so this variable was not relevant in that context.

Table 13: Results of a single regression of *dropped_out* and *declined_participation* on youth characteristics and risk level at baseline

	Dropped Out of Treatment (n=372)	Declined Participation in the Program (n=467)
Male	0.006 (-0.051)	-0.017 (-0.038)
Over 15 years old	0.087* (-0.05)	0.033 (-0.037)
Not in school	0.236*** (-0.084)	0.023 (-0.063)
4 or more risk factors	-0.088 (-0.099)	0.039 (-0.076)
Mean number of risk factors	0.025 (-0.018)	-0.009 (-0.013)
B: Weak parental monitoring	0.018 (-0.05)	0.045 (-0.037)
C: Critical life events	0.073 (-0.052)	-0.047 (-0.038)
DE: Impulsive risk taking	-0.03 (-0.062)	-0.056 (-0.045)
EG: Risky behavior	-0.007 (-0.052)	0.017 (-0.039)
F: Guilt neutralization	0.036 (-0.062)	0.037 (-0.047)
G: Negative peer influence	0.055 (-0.051)	0.046 (-0.038)
H: Peer delinquency	0.032 (-0.05)	0.008 (-0.037)
IJ: Self-reported delinquency	0.023 (-0.057)	-0.118*** (-0.041)
T: Family anti-social influence	0.022 (-0.05)	-0.027 (-0.037)

Note: Coefficient, standard error and level of statistical significance reported. * p<0.10, ** p<0.05, *** p<0.01

CAREGIVER SURVEY ATTRITION

82 caregivers who participated in the survey at baseline failed to participate in the survey at endline. We determine if caregivers who dropped out of the survey are different than caregivers who remained in the survey. Table 14 shows the results of the single regression of caregiver and youth characteristics and outcomes at baseline on survey attrition.

Differences between caregivers who completed the survey at endline and those who did not are mostly small and not statistically significant. Caregivers who dropped out of the survey had higher levels of education than those who remained in the survey. This difference is statistically significant at the 95 percent level. Across dimensions of all other caregiver characteristics, youth characteristics and youth outcomes of interest, caregivers who completed the endline survey are largely the same as those who did not complete the survey. Thus, caregivers dropping out of the survey likely did not impact study results in any meaningful way.

Table 14: Results of a single regression of survey attrition on caregiver and youth characteristics and risk level at baseline

	Caregivers who dropped out of the survey
Caregiver characteristics	**(n=940)**
Parental status	-0.0446 (0.024)
Male	0.0242 (0.028)
Age	-0.00124 (0.0009)
Education	0.0303** (0.011)
Not employed	0.0282 (0.019)
Youth characteristics	**(n=883)**
Age	.0015 (.003)
Male	-0.0232 (0.0128)
Out of school	-0.00200 (0.0211)
Overall risk level (number of risk factors)	-0.00130 (0.00450)
A: Anti-social tendencies (mean)	-0.00179 (0.0120)
B: Weak parental monitoring (mean)	-0.0103 (0.00796)
C: Critical life events (count)	0.00503 (0.00448)
DE: Impulsive risk taking (mean)	-0.00163 (0.00921)
EG: Risky group behaviors (count)	0.000151 (0.00312)
F: Guilt neutralization (mean)	-0.00714 (0.0122)
FSH: Horizontal family (mean)	0.000209 (0.0103)
FSV: Vertical family (mean)	0.00389 (0.00733)
G: Negative peer influence (mean)	0.00511 (0.00708)
H: Peer delinquency (mean)	0.0125 (0.00988)
IJ: Self-reported delinquency (count)	-0.00276 (0.00195)
T: Family antisocial influence (sum)	0.000440 (0.00249)

CAREGIVER PROGRAM ATTRITION

We explore if caregivers in the treatment group who dropped out of the program or declined to participate after being randomized to treatment are different than participants who completed the program. Table 15 shows the relationship between participation status and characteristics and outcomes of interest. Column 1 reports a coefficient and standard error of a simple regression of dropping out of the program on caregiver and youth characteristics and risk factors at baseline. Column 2 reports a coefficient and standard error of a simple regression of declining to participate on caregiver and youth characteristics and risk factors at baseline.

There are no statistically significant differences between caregivers who dropped out or declined the program across caregiver and youth characteristics. Caregivers who declined to participate in the program are more likely to have youth who scored lower on module IJ: Self-reported delinquency than youth of caregivers who took up the treatment.

Table 15: Program attrition on caregiver characteristics, youth characteristics and youth risk level at baseline

	Dropped Out of Treatment	Declined to Participate in Treatment
Caregiver characteristics	**(n=319)**	**(n=360)**
Parental status	-.031 (.066)	.079 (.051)
Male	-.052 (.081)	.0225 (.0651)
Age	-.000 (.003)	-.003 (.002)
Education	.004 (.032)	.013 (.025)
Not employed	-.044 (.057)	.007 (.044)
Youth characteristics	**(n=307)**	**(n=369)**
Age	.017 (.0137)	.000 (.010)
Male	0.00623 (0.0534)	-0.00150 (0.0398)
Out of school	0.180 (0.0934)	0.0238 (0.0690)
Overall risk level (number of risk factors)	0.0193 (0.0183)	-0.0145 (0.0140)
A: Anti-social tendencies (mean)	-0.00539 (0.0479)	-0.0243 (0.0359)
B: Weak parental monitoring (mean)	-0.00629 (0.0328)	-0.0161 (0.0247)
C: Critical life events (count)	0.00780 (0.0184)	-0.0186 (0.0134)
DE: Impulsive risk taking (mean)	-0.0220 (0.0354)	-0.0268 (0.0274)
EG: Risky group behaviors (count)	0.0163 (0.0128)	0.00419 (0.00945)
F: Guilt neutralization (mean)	0.0253 (0.0508)	0.00949 (0.0391)
FSH: Horizontal family (mean)	0.0373 (0.0417)	-0.00194 (0.0318)
FSV: Vertical family (mean)	0.0498 (0.0303)	-0.0213 (0.0232)
G: Negative peer influence (mean)	0.0172 (0.0307)	0.00894 (0.0229)
H: Peer delinquency (mean)	0.0557 (0.0389)	-0.000638 (0.0296)
IJ: Self-reported delinquency (count)	0.00298 (0.00852)	-0.0202** (0.00622)
T: Family antisocial influence (sum)	0.00108 (0.0110)	-0.000370 (0.00814)

DISCUSSION OF THE TREATMENT AND CONTROL IMPROVEMENTS

To explore the reasons that could drive control group improvements in the YSET outcomes, we investigated several potential explanations. In this section we summarize a literature review conducted on surveying youth and the impact that surveys may have on behavior. Second, we analyze CFYR's community survey to check for community-level changes. Third, we present our analysis of potential spillover effects. Finally, we discuss some potential shortcomings of the YSET as an evaluation tool.

Findings from qualitative research in the three countries help to inform and position the evaluation results. In each country, we conducted focus group discussions with caregivers and youth who were participating in the program and held meetings with family counselors, CFYR staff, and other key stakeholders. In addition, we base many of our insights in a theoretical account through a decision tree of the phenomena that can undermine the measurement of risk differences between treatment and control groups. The theoretical discussion is included in ANNEX VI: THEORETICAL FRAMEWORK REGARDING OBSERVABILITY AND IDENTIFYING CAUSAL EFFECTS, but it highlights the way in which targeting of interventions is connected to the intrinsic difficulty of accurately measuring risk. In addition, the theoretical discussion includes causal graphs to highlight the challenge of causal identification that underlies the whole impact evaluation.

YSET TIMESTAMP DATA ANALYSIS

We examined timestamp data on the YSET at baseline, midline, and endline to check how long it took respondents to complete the survey, and if there were mean duration differences by age, school status, country, gender, treatment status, program participation status, program completion status, or youth risk level. After removing outliers who took less than 10 minutes or more than three hours to complete the survey, we found that, on average, respondents took 38 minutes to complete the survey. A t-test of differences in means revealed that there were statistically significant differences in three of the categories listed above: age group, youth risk level, and country. Younger respondents aged 10 to 14, on average, had shorter survey durations (37.2 minutes) than older youth aged 15 and above (38.4 minutes). We had expected to find that younger youth took longer to respond because of greater challenges comprehending the questions. The higher mean duration for older youth may in part be explained by their higher risk levels, since for some of the modules, when youth answer yes to a risky behavior, the survey opens up additional questions. Lower-risk youth who scored a 4 or a 5 on the YSET on average completed the survey in 37.2 minutes, and higher risk youth who scored between a 6 and a 9 had an average duration of 42.6 minutes. Respondents in Guyana on average had the shortest duration (35.4 minutes), and respondents in St. Lucia had the longest (42 minutes).

LITERATURE ON YOUTH ANSWERING SURVEYS

It is not easy for youth to answer a questionnaire like the YSET. Questionnaire fatigue and the respondent's level of comprehension of YSET questions are both concerns that we explored during the qualitative fieldwork trips. In our discussions, youth and enumerators alike mentioned that the questionnaire felt long, and that there were challenges with keeping attention as respondents or explaining

the questions in the case of enumerators (to ensure youth were answering thoughtfully and attentively). In this section, we present the findings from a literature review exploring these issues, and particularly trying to explain the control group improvements in YSET outcomes.

One possible avenue through which the control group may have changed could be related to the changing behaviors, attitudes, beliefs and cognitive capacities of the youth in the year that passed between baseline and endline.[16] The way in which youth approached the survey instruments may have differed at midline or endline, as compared to baseline. This is particularly important for the control group, since the changes they experienced cannot be credibly explained by the intervention.[17] If there is a specific reason for the control group to behave in more extreme ways than the treatment group when answering to survey instruments, this may undermine the possibility of detecting an effect from the program.

Although the YSET was pretested and adapted to the specific conditions of the Caribbean, it is possible that youth approached questionnaires in ways that may have been detrimental to the reliability of this instrument as an indicator of behavioral and attitudinal changes. If those reliability issues were more prevalent among control youth, we may be able to make sense of the large change experienced in that group. Such changes may offset program effects that would have been more evident had the control youth not answered the YSET differently at endline. As it turns out, most of the potential explanations related to survey design or response appear to apply equally to both control and treatment youth. They do not seem to account for the patterns of improvement observed in control youth, even though they may explain, to some extent, why all the youth, regardless of treatment, community, age group, gender or country, shifted in their YSET scores to lower risk through time.

The literature on survey design provides insight into the types of challenges researchers face when interpreting results from surveys applied to youth. The problems range from the cognitive capacity of younger respondents who may simply not quite understand what is being asked; to the possibility of differences in the propensity to satisfice, namely making the least possible effort when answering in order to get the questionnaire out of the way; to issues of questionnaire fatigue, question order or a propensity to lie or provide the socially desirable answers; and even potential Hawthorne effects arising simply from the attention researchers are placing on the youth.[18] Many of these issues are obviously high in the mind of the team that designed the YSET in order to provide a reliable measurement instrument; however, it is worth discussing how they may have impacted the particular case of CFYR. As will become clear, the literature suggests that there may be some age-specific characteristics that may enable youth who were younger when they were first exposed to the YSET to provide systematically different answers when they are a year or two older, but none of them suggests a differential effect between treatment and control youth.

Younger children may face greater cognitive challenges in a face-to-face interview, which may affect the quality of collected data (Fuchs, 2005 and 2009). However, this is likely to influence the non-response rates or the way in which they deal with numeric scales, not make them more likely to report a type of

[16] Hurricane Irma happened in September 2017, right when data collection for the baseline YSET had begun. This caused some delays in data collection and program start dates. Annex II has a table with all dates for data collection and program implementation.

[17] Even if there were spillover effects, it is not conceivable that those effects would be as strong among the control youth as the treatment effects for treated youth and their families.

[18] The Hawthorne or observer effect refers to the possibility of individuals to reactively modify their behavior as a consequence of their awareness of being observed. Social desirability bias refers to the tendency to respond questions in a manner that will be viewed favorably by others.

behavior or attitude over another. Although youth become one to two years older as they progress to the endline, differences within each survey by an additional year or two of age are not large enough to explain the sharp reduction in risk observed over the course of the evaluation. Youth reported in midline focus group discussions that they did recall responding to the YSET at least twice. Whenever we asked about the YSET, we were extremely careful to never mention the instrument by name, nor did we provide any indication that it may be important, in order to ensure the integrity of the use of the instrument.

Being in school, children are quite familiar with a format of testing and they may interpret a survey as a test in which they are expected to provide "correct" answers to questions. This seems to be more prevalent when a survey is collected in schools (Scott, 2000), but one should recall that a large number of surveys, particularly for some of the higher-risk youth not in school, were collected at home. Treating the survey as a test would not explain any particular effect for the control group. The literature suggests that older youth may manipulate the instruments by purposefully answering "as a joke," but this behavior may be countervailed by a greater nuance in their sense of morality and their desire to conform with peer pressure (De Leeuw, 2011). These behaviors may exist, but they do not account for a systematic shift in the control group as distinct from the treatment group.

Youth reported in one of our focus groups at midline that they would sometimes choose a different answer because they remembered from baseline that answering "YES" would open up a whole set of additional questions, only to discover that "NO" also opened up additional questions. Older youth seemed to understand that they could answer the questionnaire strategically, but they also overwhelmingly told us that they responded relatively truthfully. Younger children may be more likely to use satisficing when answering questions on subjects they find uninteresting (Borges et al., 2000). In our findings, there are some statistically significant heterogeneous age effects for a couple of YSET scales. However, we find no systematic evidence of differential effects by age or for the YSET threshold. Even if there were differential propensities to answer some questions with lower risk as a youth becomes older, this does not suggest an avenue that would motivate a different behavior for control group youth as compared to a youth in the treatment group: in both groups, youth are aging, maturing, and crossing the threshold of 18, on average, at the same pace.

It is possible that once enrolled in the program, some youth would suffer from social desirability bias (Atkins-Burnett, 2016), in terms of wanting to show that they had shifted their behavior and attitudes toward lowering their risks. Social desirability bias is a common problem in social science research. Surveys referencing sensitive topics in particular may illicit results distorted by social desirability bias (Krumpal 2013). This could be especially true for youth, as social desirability has been found to account for variance in self-reported deviant behavior among youth (Camerini and Shulz 2018). However, if this were the most salient motivation behind the reduction in risks, one would expect that the behaviors the youth had been working on most with the counselors, which were scales A, B, FSH and FSV of the YSET, would be the ones where we would observe the largest shifts. In other words, in order to please the counselors, youth would deliberately show improvements in areas where they worked with counselors the most. Looking at these four scales in particular, and the YSET as a whole, we do not see evidence of social desirability bias in how youth answered the survey. One common way in which social desirability bias manifests itself is through non-responses on certain types of questions, and the share of non-responses in our YSET sample is extremely low, with zero or very close to zero in all questions.

Although social desirability bias may also exist among the control group of youth, the motivation to please the enumerators or the program might be less prevalent among them. In fact, there might even be a countervailing motivation, in that if youth are trying to remain eligible for the program, they may find it useful to report antisocial or other less desirable behavior. Males may even be willing to lie, in order to pretend to be bigger risk takers than they really are. However, we found no evidence of such behavior in the qualitative fieldwork, nor would this be likely to impact the control group differentially from the treatment group.

Survey length can also influence the way youth respond to questions. Evidence of survey fatigue includes respondents answering questions faster at the end of questionnaires and providing more identical answers to multiple questions in a row (Galesix and Bosnjak 2009; Herzog and Bachman 1981). Youth, specifically, have been found to fabricate answers when responding to longer instruments (Betrand et. Al 2009). YSET survey length has been a concern for researchers since the instrument was implemented in Los Angeles (Kraus et al, 2017).

There is a final possible influence in the YSET surveys that may change behavior due to the scrutiny that comes from having to answer complex surveys with relative frequency (three times over the course of one year), but such an effect would impact both the control and treatment groups. This could be considered a "substantive Hawthorne effect" (Murray, 1988), in which intense questioning regarding risky behaviors may actually make youth more aware of their behavior and trigger a desire to move away from undesirable and antisocial attitudes and behaviors. While this may be at work in the shift in the risk reduction in the YSET scales, it is unlikely that it would generate larger movements than those generated by the intervention itself.

The comparative literature on survey instruments collected from youth respondents does not provide much insight into why the control group would uniquely improve during the course of the evaluation. All the potential effects related to questionnaire design, social desirability bias, cognitive maturing, recall of the previous time the instrument was answered, or a strategic manipulation of responses seem to apply equally to control and treatment youth. The only possibility that may affect them differentially is the substantive Hawthorne effect, in that control youth may be triggered to reduce risky behaviors when reminded by the questionnaire that those behaviors are undesirable. Perhaps this awareness was reinforced by the prospect of becoming program beneficiaries in the future. Even if these effects existed, however, there is an equal chance that they would also be found, even if in attenuated form, among the treatment group.

SPILLOVER EFFECTS

Spillovers are indirect effects that the program may have had on people who are not receiving the treatment, such as control youth. Spillovers are common for interventions such as medications that prevent or cure contagious diseases, because those who get treated will not be infecting others (for an example of a deworming intervention with spillover effects, see Miguel and Kramer, 2003). In the case of a family counseling intervention, a similar argument applies. If the treatment youth's behavior is changing outside the home, such as fighting and engaging in delinquent behavior less often, control youth who would be co-participating in these behaviors would also be fighting and engaging in delinquent behavior less often. In addition, one can imagine that youth and caregivers may discuss the counseling and what they are learning with others outside the program. Given the unexpected drop in risk detected by the YSET among

control youth, one possible explanation is that the benefits of the program spillover to treatment youth. In this section we explore this possibility.

QUALITATIVE FIELDWORK

As part of our midline qualitative research trips to the three countries, we asked youth and caregivers during focus group discussions if they had discussed the *Family Matters* program or what they were learning with anyone outside their household. The vast majority of focus group participants told us that they had not discussed it with anybody. Neighbors might see the counselor coming and going, but participants reported that people tend to mind their own business, and if neighbors did ask what the visits were about, participants told us they would not share.

Anecdotes shared by participants suggested that the counseling primarily improved home behaviors such as arguing, yelling, giving attitude, doing chores, and the child letting the adults know when he/she is going out. Some youth mentioned better school performance. These behavioral changes revealed during qualitative research are not behavioral changes that would necessarily cause improvements in control youth's risk factors as well, because the improvements are contained to the home and school grades.

We noticed some selection bias in focus group discussion attendance, where the *Family Matters* participants that were excited about the program were more likely to participate. Our field coordinators invited a random selection of program participants to focus group discussions, and sometimes the participant at the other end of the phone line expressed dissatisfaction with the program. Those participants who seemed to be more critical, despite much effort on the part of the field coordinators to get them to attend, did not show up. This selection bias should, if anything, make the possibility of spillover effects become more evident in the discussions because the most excited participants would be more likely to change their behaviors and talk to others about it. However, we still did not find any evidence of spillover effects based on our qualitative research.

YSET ENDLINE SURVEY EVIDENCE RELATED TO SPILLOVERS

In some experiments, there are ways to design the randomization procedure to measure spillovers. However, the country contexts would have made it unfeasible to design this experiment to measure spillovers. A small sample size due to small population sizes in each community confounded our ability to configure randomization in a way to quantitatively calculate spillovers. Instead, our best account of spillover effects is captured through questions implemented at the end of the program. At endline, we asked treatment youth whether they discussed the program with others. Less than half of treatment youth (48 percent) said they had discussed the program with their friends or neighbors. When asked whether they have any friends who are also enrolled in the program, 58 percent said none, 22 percent said one, 9.5 percent said two, and 11 percent said three or more.

At endline, control youth were also asked a few questions about the *Family Matters* program. A total of 92 control youth (18 percent) had heard of the program (see

Table 16). Of the control youth who had heard of it, only 42 percent knew somebody who was participating (8 percent of the total). Therefore, most of the control group youth had not heard of the program and did not know anyone who was enrolled.

Table 16: Questions about *Family Matters* to control youth

	Yes	No	Don't Know
Have you heard of the Family Matters program? (n=498)	18%	82%	-
Do you know any youth participating in this program? (n=92)	42%	58%	-
Do you think this person is now less likely to get in trouble at school or at home, because of the program? (n=39)	46%	41%	13%

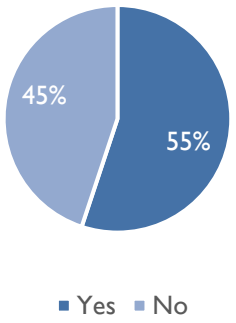
Armed with information about which control youth know someone enrolled in the *Family Matters* program, we checked if these control youth are the ones driving control group improvement in the YSET. We ran a simple linear regression of the outcome "change in the number of YSET risk factors" on the binary variable of knowing a *Family Matters* program participant. There is not a statistically significant difference in overall risk reduction between control youth who know or do not know program participants.

We also ran the difference-in-difference analysis with the four models on the TOT sample excluding the control youth who know somebody in the program. We wanted to see if there is a statistically significant difference between treatment and control groups when we exclude the control youth who may have benefited from spillovers. The coefficients for the outcome "mean number of risk factors" remained virtually unchanged, at -0.430, and retained the same level of statistical significance. Therefore, the control youth who know somebody in the program are not any different in terms of changes in YSET risk levels and are not driving the analysis results.

CAREGIVER ENDLINE SURVEY EVIDENCE RELATED TO SPILLOVERS

To further explore the potential for spillovers, we asked treatment caregivers at endline if they discussed the program with others. Figure 43 presents results from this question. The majority of treatment caregivers did discuss the program with others. Obviously, we do not know the nature of these conversations. However, if caregivers shared advice or tools for communicating with their children, individuals outside of the program could have benefitted from such knowledge.

Figure 43: Treatment caregivers discussing program with others



It is important to note, however, that we do not see the same pattern of dramatic change in both treatment and control groups in the caregiver survey. Responses for both groups are fairly stable between baseline and

endline. This, combined with the lack of evidence of spillovers in the YSET analysis above, suggests that spillovers are not likely the cause of the parallel changes in YSET control and treatment responses.

CONFOUNDING FACTORS

COMMUNITY CHANGES

Communities may have changed in ways that would systematically shift risky behavior and attitudes in both treatment and control youth. Since community surveys were collected at baseline and a later period during the course of intervention (either at midline or endline), it is possible to use the information from the aggregate conditions of the community to find out whether the movement in the control group may be attributed to a change at the community level. This is not the same as measuring a spillover effect, since the shift may be caused by unrelated confounding factors, such as changes in economic activity, employment prospects or the environment where youth and their families carry out their day-to-day activities.

During the course of the evaluation, two major events can be identified that may have provoked some changes across the board in all communities in the three countries, although they are unlikely to affect all of them in the same way. In Guyana, a major discovery of oil and natural gas potential with the Stabroek Block's Liza oil field (now in operation), as well as 16 more oil discoveries between 2017 and 2019,[19] generated enormous optimism regarding the prospects for growth and development in that country. Hurricane Irma, on the other hand, impacted St. Kitts Nevis quite dramatically in September 2017, although the catastrophic damage of the storm was not as severe as in the other Leeward islands. The recovery phase may have reignited some economic growth, although it is important to note that economic dynamism with the exception of Guyana is relatively low, with a lackluster average growth rate for the whole region, according to the Caribbean Development Bank, of 1.6 percent in 2018 and 1.0 percent in 2019.[20] Hence, it is unlikely that country-level events or effects are driving the change in risk observed during the course of the evaluation.

The community survey data collected by CFYR allows us to compare victimization rates and the perceptions of how prevalent criminal activity was in each country and community. Victimization is measured in the survey as the share of respondents that report having been victims of crime during the prior year. Perceptions on criminal activity are summarized into an index, using polychoric factor analysis, from feelings of insecurity and vulnerability regarding robbery, sexual assault, safety walking after dark, vandalism, drugs and overall feelings of safety and crime levels (details on the indices are provided in ANNEX IV: OUTCOME VARIABLE CONSTRUCTION At the most general level, there is no statistically significant difference in the victimization and the perceptions of criminal activity from one data collection from one period to the next. However, the average hides important differences between countries and communities that are statistically significant. Communities that have improved in these indicators are offset by communities that have become worse, leaving the average unchanged.

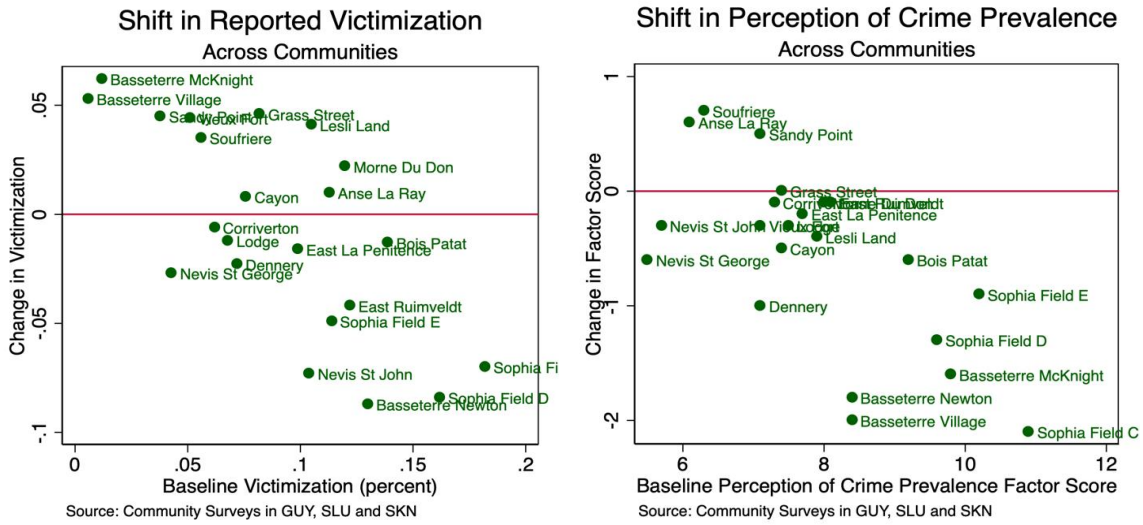
Such differences are visually summarized in Figure 44 which shows a scatterplot of community victimization changes. The horizontal axis is the average share of residents being victimized in each CFYR community at baseline. The vertical axis shows the change in the factor score from baseline to the second time when community surveys were collected (near the intervention completion date). Any community above the zero horizontal line represents an increase in victimization, which ranges from a six percent increase to a 10 percent

[19] Exxon Mobil. 2020. Guyana Project Overview.
[20] Caribbean Development Bank. 2019 Caribbean Economic Review and 2020 Outlook.

decrease. Although there are more communities below the line, there is no obvious shift in reduced victimization across the board. The alignment of the communities along a descending diagonal suggests a certain degree of convergence, in which the communities that used to have lower victimization have increased their percentage, while the communities that had the higher victimization have reduced it. However, the communities are quite spread out. For example, at the level of 10 percent victimization, communities experienced practically the full range of potential victimization changes in the period we measured.

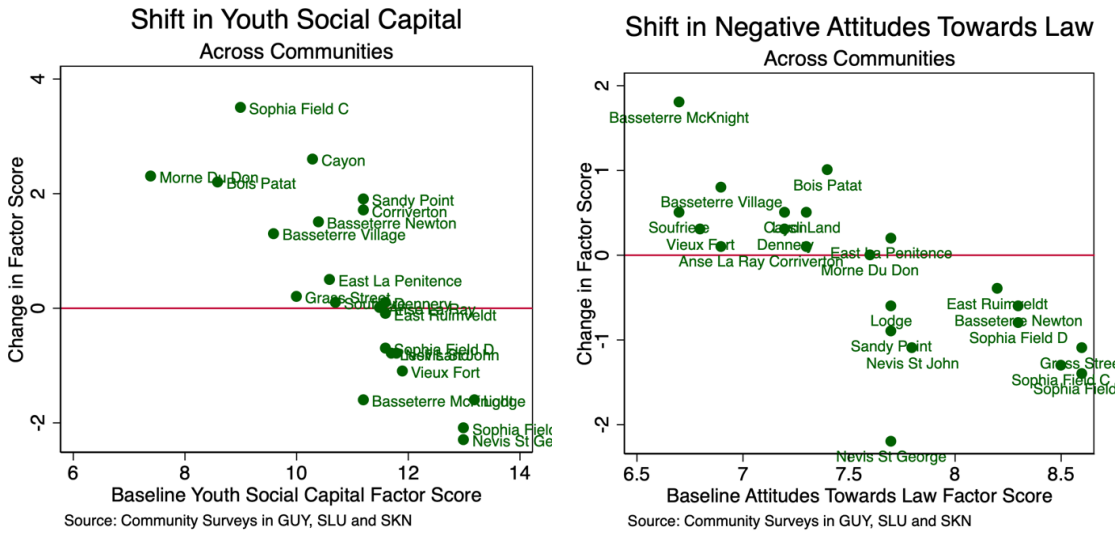
A similar pattern emerges from the evolution of the crime prevalence perception factor score, although in this case the metric is an index. Most communities have changed very little, concentrated at around the value of eight and zero change. The largest reductions in crime perception happened in communities in Georgetown in Guyana, and Basseterre in St. Kitts and Nevis. This decrease in crime perception in Basseterre matches what we heard during the midline research trip, where locals told us that the local gangs were currently "at peace" and tensions and violence had decreased. Increases in perception are observed in communities outside the capital city in both St. Lucia and St. Kitts and Nevis.

Figure 44: Shifts in reported victimization and perception of crime prevalence by community



A similar analysis can be carried out for one of the measures of social capital, which is specific for youth, and the negative attitudes toward the law as expressed by the residents of those communities. In the case of the social capital index, which is only calculated for youth 10 to 17 interviewed in their communities, it includes feelings of inclusion, voice and membership in the neighborhood, perceived good influences, bonds and a sense of belonging to the community. The index regarding attitudes toward the law is coded negatively, in that community members are asked whether the law does not protect their interests, whether it is ok to break the law, their feelings of distrust in the legal system, and police corruption. In Figure 45 we observe that the pattern of convergence is similar in these indices, but there is no clear pattern regarding specific countries having experienced larger improvements or declines in these metrics.

Figure 45: Shift in social capital and negative attitudes towards the law



Hence, the community survey does not suggest that risk reduction in the YSET can be attributed to community-level confounds. Communities experienced the whole range of changing conditions, and if there is any pattern, it is a trend for convergence in the same kind of experiences being shared by communities across the three countries.

OTHER PREVENTION PROGRAMS

Another possible cause for risk reduction among youth is participation in other activities and programs. To examine this possibility, all youth surveyed at endline were asked about their after-school and weekend activities in the last year. 40 percent of youth reported having held a paying job in the last year, with 51 percent of youth aged 15 years and above having held a paying job, and 17 percent of those aged 14 and under having held a paying job. Treatment youth were more likely than control youth to be in school and working during the prior year.

Table 17: Share of youth who held a paid job in the last year

	All Youth (n=980)	Youth Who Completed the Treatment (n=230)	Control Youth (n=499)
Had a paid job last year	40%	42%	39%
Mean number of work hours in a typical week	30	28	30
Out of school and held a paid job	66%	65%	64%
In school and held a paid job	28%	34%	26%

Table 18 shows a breakdown by the kind of activity that youth reported participating in during the weekday and weekend. To check whether youth who report taking part in these activities experienced different levels of risk change, we ran a simple linear regression of the outcome "change in the number of YSET risk factors" on the binary variable of engaging on a weekday activity, and separately another simple linear regression of the outcome "change in the number of YSET risk factors" on the binary variable of engaging

on a weekend activity. We did not detect a statistically significant difference in overall risk reduction between youth who participated in these programs in the last year or not.

Table 18: Share of youth who spent time on weekday and weekend activities

	All Youth (n=980)	Youth Who Completed Treatment (n=230)	Control Youth (n=499)
Weekday Activities			
None	38%	30%	37%
Sports	38%	44%	38%
Music/Dance/Art/Cultural	13%	16%	16%
Professional or Educational Training	15%	17%	17%
Church/Temple/Mosque/Other Place of Worship	12%	12%	13%
Community-Related Activities	6%	9%	5%
Volunteer Work	5%	8%	5%
Other	3%	5%	3%
Weekend Activities			
None	50%	45%	49%
Sports	28%	30%	28%
Music/Dance/Art/Cultural	10%	13%	9%
Professional or Educational Training	7%	8%	7%
Church/Temple/Mosque/Other Place of Worship	17%	19%	19%
Community-Related Activities	4%	6%	4%
Volunteer Work	3%	6%	2%
Other	4%	6%	4%

In summary, the different sources of information that we investigated to better understand why the control group shows improvements in the YSET do not point to spillovers or other known effects. The most likely explanation then is that treatment and control youth's risk level is similar at endline, and the drop in risk for both groups concerns the way they relate to the YSET as a (long) instrument that gets repeated. The different effects that could cause youth to lose interest, satisfice, or give "correct" answers are expected to influence control and treatment youth in a very similar or equal manner.

CONCLUSIONS

PROGRAM EFFECTIVENESS

The pre-analysis plan for the evaluation clearly established expectations regarding the kind of indicators that would provide a sense of program impact. We specifically established that the YSET and its components would be used to measure the program success. In addition, we committed to exploring potential effects on caregivers. On the technical front, our design sought to ensure that the evaluation had enough statistical power in order to detect even small program effects. The pre-announced methodology for the analysis of the data ensured that the evaluation team would measure impact with the best analytic tools available. The basic threshold of a positive evaluation of the *Family Matters* program would require finding program effects that are statistically significant and suggest a substantial change in risk.

The CFYR intervention was meant to impact youth and their families by reducing risk factors and making children and youth less likely to engage in anti-social or risky behavior. The program implementors are committed to improving the conditions for youth and their families in the Caribbean. They are professionals who possess the expertise to train counselors, oversee program implementation and carry out complex activities often requiring swift adaptability to conditions on the ground, and above all, very hard work.

From what we could tell during the qualitative research trips, counselors were engaging families according to the procedures established by the program and adhered to the manual. They seem to have been trained well to follow established protocols but were also able to adapt the work to the reality on the ground. In Corriverton, for example, counselors found that illiteracy in some families prevented them from being able to complete some of the activities, so the counselors created a handout drawn by hand to visually explain the activity.

An impact evaluation cannot gauge staff motivations, levels of effort or the specific procedures put in place by CFYR to carry out the program. One hypothesis for a program not having strong effects is a problem of implementation. Observation of the family counselor training before the program started and midline qualitative research trips to the three countries suggested that poor implementation is not one of the reasons here. The observation and interviews done by the SI team attest to the solid commitment and professionalism of the program implementer, one that used its international experience and good judgement to put in place the best program they could offer. Those trips provided our evaluation team with opportunities to talk to the youth and the caregivers to learn, articulated in their own words, about how the program impacted them. We remain deeply impressed by the stories and insights of how families enthusiastically told us about what a difference the program and individual counselors were making in their everyday lives. We reiterate how grateful we are to Creative Associates and CFYR in opening up the opportunities and spaces necessary for carrying out a randomized evaluation, explaining the various elements of the intervention and cooperating with us to ensure a process of robust data collection.

During the midline qualitative fieldwork, we held discussions with family counselors in the three countries. Counselors could see the positive impact of their work and believed in the family counseling strategy. They had some suggestions for changes to increase program effectiveness: 1) for counselors to be able to work on more than one issue at a time (as opposed to only one issue, as stated in the manual), 2) despite

the great benefits they saw of holding the family sessions at home, to have the option for individual (youth) counseling to take place outside the home because of privacy concerns, 3) to be allowed to double check some areas of the intervention, for example, school attendance, 4) to be provided with money to help families meet basic needs (for instance, one youth stays out late because he gets food from friends, and some families will also need assistance getting started with activities), 5) to have their work supplemented by a parental component (with parenting guidance), and 6) to be able to provide specific feedback directly to the implementers rather than use the Model Fidelity Database (MFD) platform to record the work due to problems with the platform.

Youth and caregiver program assessment and focus group discussions revealed high levels of participant satisfaction with the program. Caregivers, in particular, described changes that they have seen in their children and in themselves because of the work of family counselors. Most commonly, they described better communication by the family and improved temper or attitude by the youth. Youth were a bit less enthusiastic, both in the program assessment part of the survey and in focus group discussions, but still most of them embraced and recommended the program.

During our qualitative research trips, we heard anecdotes of how the program has personally changed the participants. One single mother in Guyana shared that she now conducts weekly family counseling sessions with her own children. She asks them to come talk to her one at a time, about something that is bothering them or anything else they would like to talk about. Next, the mother takes time by herself to reflect on what she heard, and then shares her thoughts with each of them. This mother had clearly been positively influenced by the work of the counselors and saw the benefit of improving communication within her family.

We do not doubt that the intervention had some positive impacts on some participating households, but the impact evaluation results suggests that these impacts were either not adequately widespread or that similar improvements occurred among control families for other reasons. Empirically, we can only say that the program has had a very small effect on the risk levels of youth, and no measurable impact on the caregivers. The evaluators tested for changes between baseline and endline using four difference-in-differences models, and used three different samples: ITT, TOT, and a matched TOT sample. The results are consistent across models and across samples.

Over the period of the evaluation, in all three countries, the youth studied experienced a substantial decrease in risk factors, as measured by the YSET. This was true both for the treatment as well as the control group as determined by the randomization.[21] During the year of the program's duration, the YSET mean number of risk factors decreased for the control group of youth, that did not receive the intervention, to the point that half of them are no longer eligible for the program. The average number of risk factors for the control group dropped from 5.1 at baseline to 3.8 at endline (representing a drop of 1.2 factors). Rather large decreases in risk were observed in all three countries, and in every YSET scale for both treatment and control.

[21] As discussed in the literature review, the impact evaluation of the *Proponte Más* program in Honduras found similar evidence of large control group improvements throughout the YSET. The outcome "four or more risk factors", which determines eligibility for the program, did not detect statistically significant differences between treatment and control after the intervention. The most recent evaluation of the Los Angeles GRYD program also found statistically significant changes in risk in the control group they used for comparison. A larger endline difference between treatment and control was detected in the Los Angeles case (and we note concerns about their control group in the literature review), but the most important point here is that all three evaluations found large reductions of risk in a group that did not receive the intervention. We believe these findings support our conclusion that the YSET needs to be closely examined.

Nonetheless, on some indicators, the evaluation found a statistically significant change in the treatment youth, compared to the control group. Specifically, 9.3 percent more treated youth were below the risk threshold of four or more risk factors than control youth, and the drop in the mean number of risk factors was larger by 0.42 points (in the zero to nine YSET scale) in the treatment versus control group. On average, treatment youth experienced a 34 percent reduction in their mean number of risk factors from baseline to endline (from 5.12 to 3.37), and control youth experienced a 26 percent reduction (from 5.09 to 3.75). While statistically significant, these are rather modest effects, at least compared to the changes experienced by the youth over time. When the number of risk factors is standardized by dividing it by the standard deviation it gives an effect size of 0.189, or a fifth of a standard deviation. In the evaluation literature, this is considered to be a small effect size. An alternative estimation approach that matches a smaller control group more closely to the treatment group finds an even smaller and not statistically significant relationship.

Of the nine individual scales, only two of them show a statistically significant difference, and again the relative size of the effects is rather small. Scale F (Guilt neutralization) shows a consistent robustness to the way it is measured; in that it is also statistically significant when the mean instead of a dichotomic threshold is used. However, the results for scales T (Family antisocial influence) and DE (Impulsive risk taking) are too fragile to survive a different way of measuring them. It is important to keep in mind that just by sheer probability, one is likely to find that one out of 20 coefficients of a given estimation may turn out to be significant at a 95 percent level. Thus, we are not confident that the evaluation is finding robust evidence of how the intervention affected the behavior or attitudes of youth. Moreover, we fail to find statistically significant effects in the scales that the counselors appeared to have been concentrating a substantial amount of their effort in changing, as revealed by the midline qualitative interviews we had with them. These were scales A (Anti-social tendencies), B (Weak parental monitoring), FSH (Family scale – horizontal) and FSV (Family scale – vertical).

The statistically significant results in the "mean number of risk factors" and "four or more risk factors" outcomes are primarily driven by Guyana. In St. Kitts and Nevis, the average scores of youth in the treatment group actually increased slightly from baseline to endline, and in St. Lucia, there is not a statistically significant difference between treatment and control changes. Among caregivers, there is no clear pattern in the way in which the treatment and control groups change over time. In some indices, both groups improved, in others there was virtually no change at all, and yet in others one group improved a bit more than the other. No indicators show statistically significant results. Thus, the empirical evidence fails to prove that caregivers' management of their households and experiences with youth are improved by the *Family Matters* program. In summary, looking across the many outcome indicators examined both through the YSET and the caregiver survey and the many estimation approaches, we conclude that the intervention may have had a minor impact on overall risk, but that overall the intervention was not effective in lowering youth risk and improving family-level outcomes. We do not find evidence of spillover effects, confounding factors, or weak implementation that would help explain these findings. We do, however, raise some concerns with using lengthy assessment tools to measure change in youth risk over time.

LOOKING AHEAD

The impact evaluation teaches lessons that cannot be obtained from a simpler assessment of client satisfaction or from an observational study tracking youth through time without a control group comparison. Robust impact evaluations are the only way to know whether a program is having an effect, by making a comparison to a true counterfactual of no intervention. This is the reason why program impact evaluation must remain a critical aspect of how USAID works to improve its programming in the region.

Some of the recommendations we make are more likely to be relevant for governments, others may be useful to USAID seeking to support similar programs in the region, and others may be important for agencies and partners doing implementation on the ground.

1. The evaluation has raised concerns with the YSET as both an evaluation and assessment tool. There is some evidence to suggest that the instrument is long, youth were not engaged throughout the assessment, and youth may have had trouble understanding some of the questions. Indeed, in Los Angeles, evaluators also identified concerns about YSET "vocabulary and terminology, language comprehension for younger youth, as well as the length and number of questions" (Kraus et al, 2017). In addition, while it is reasonable to adapt a validated instrument from another context, the YSET was validated for a different purpose (to identify gang members and those at risk of joining gangs) and a different context (Los Angeles). As such, we recommend revisiting the YSET as an assessment tool and developing a revised tool. The subsequent tool should be 1) shorter, 2) tested for survey fatigue, 3) tested for producing reliable results (including tests for respondent comprehension of questions), 4) validated as a predictor of youth violence, and 5) linked to the intervention's theory of change.
2. If USAID and implementing partners seek to use an assessment tool as an evaluation tool, then we continue to recommend that it NOT be applied by counselors involved with the program, at least at midline and at endline, as this could introduce bias both through the counselor and through the youth.
3. We would imagine that future initiatives will incorporate learning from this intervention and evaluation. As such, even though we find only a minor impact on potential youth violence prevention, we do not recommend terminating family counseling-based programming and instead recommend adjusting the intervention and looking for opportunities to re-evaluate future iterations. The evaluation suggests a few potential avenues for change.
4. USAID and its implementing partners should ensure that violence prevention programs provide a clearer theory of change linking the diverse aspects of family counseling and related activities to the desired intermediate and higher-level outcomes.
5. Outcomes and evaluation tools can be more clearly linked to the activities promoted through family counselling. Many of the YSET modules may not have been directly addressed by the intervention. The survey contains 17 modules, but only nine modules are used for risk assessment. In the first six months of intervention, family counselors focused on only four of those modules, and they did not have access to the youths' answers on other modules. The remaining modules not used for risk assessment include one that asks respondents how they feel about the police, which is the clearest example of a module that is not linked to the intervention. We are proposing that, in the future, the survey tool is closely examined so that parts that are not deemed essential for risk assessment and the counselors' work be removed to shorten the survey. In addition, efforts should be made to incorporate objective outcome indicators. For example, for in-school youth, schools could provide a valuable source of data, including information on absenteeism, suspensions and other behaviors at school.
6. The family counselors' manual establishes that counselors can only work on one behavior at a time, and only on behaviors coming from the YSET scales they have access to. Family counselors reported that sometimes there were unaddressed and pressing issues (e.g., drug addiction, domestic violence, abuse, suicidal thoughts) outside of the YSET scales that prevented them from really making progress. Even though the counselors may have the training to deal with some of

these issues, they are instructed to provide referrals and not work on these themselves. As currently implemented, these issues represent a risk to the theory of change. We recommend that some experimentation and perhaps even more flexibility is given to counselors to better tailor the intervention to the specific needs of youth and families, or to work closely with referral networks to ensure that these concerns are addressed simultaneously.

7. Counselors did appear to have additional availability for treating more youth, and as such, we recommend that the burden of caseload for counselors and social workers carrying out the program be increased, even if this requires some adaptation to the way they carry out their outreach. As the intervention becomes more widely known, we expect that a lot of the hard work required to gain trust and acceptance from the youth and their families will be less burdensome, and that the counselors themselves may provide feedback on how they may be able to work more effectively, individually or in teams, to include more youth.
8. The evaluation failed to find evidence of significant changes in caregiver attitudes and perceptions. Given that caregivers could provide a valuable resource to one another, an easy modification would be to encourage interaction and mutual support between caregivers. This could multiply the work of the counselors and strengthen sustainability. During focus group discussions, we found that many caregivers liked having the opportunity to interact with other caregivers participating in the program and share experiences. In some instances, they exchanged contact information to stay in touch.
9. All prevention programs confront challenges of non-participation and attrition; however, not including the most at-risk youth represents a threat to a place-based strategy, as was employed by USAID and CFYR. We recommend that given the effort that was already made in identifying at-risk youth, a more comprehensive approach might seek to follow the so-called non-compliers with additional opportunities that may allow them to revisit their initial rejection of the program or enroll in other services.
10. In the heterogeneous effects analysis, we found no evidence suggesting that the program is more successful with some profiles of less risky youth, and there was some qualitative evidence to suggest that some treated youth did not need the intervention as much as others. Given limited resources, we recommend focusing on youth that are at the highest-risk categories and more likely to engage in violence.
11. We also recommend that USAID and its partners reassess the timing of the intervention to increase the reach of the program. On the one hand, several interviewees felt that a six-month intervention would be equally as effective and more cost and time effective. On the other hand, many caregivers hoped that the intervention would continue indefinitely. One potential alternative would be to implement an intensive six-month family counseling intervention followed by a less intensive but longer period of counseling. The goal of a shortened intervention is to enable more youth to benefit from the program without increasing the cost of the program. By shortening the program, counselors would be able to take on more youth and families over the course of one year.

ANNEX I: ELIGIBILITY DETERMINATION AND RANDOMIZATION

This annex summarizes the approach used to determine eligibility and randomize youth into treatment and control.

DETERMINATION OF ELIGIBILITY

Creative Associate's project sub-contractor USC determined youth eligibility for the intervention. The USC team added up the scores for each YSET module, identified the highest risk youth across all modules, and then calculated the average module scores across this highest risk group. These averages were then used in conjunction with the distribution of the remaining youth to inform a purposively selected cut-point for each module on a country-by-country basis. Using Guyana as an example, Table 19 provides the average risk scores for the 34 highest risk youths identified and the final thresholds. The thresholds are somewhat different across each country, and as such, raw scores cannot be compared across countries.

Table 19: Average scores for 34 highest risk youth and final thresholds across nine risk modules in Guyana

	B	C	DE	F	G	H	T	IJ	EG
Average scores for highest risk youth	2.21	3.97	3.45	2.92	2.24	2.11	6.24	5.91	4.47
Final cut-points	2.5	4.0	3.0	2.5	2.3	1.9	6.0	3	3

This approach is similar to that used by USC in Los Angeles and Central America, although in these locations the scores of gang members were used to identify the thresholds. In the Caribbean, where gang membership is far less common, this approach could not be used.

MATCHING AND RANDOMIZATION

Given that the intervention targets families, one of the challenges with randomization was ensuring that two eligible youth in the same household were randomized to the same treatment condition. After matching youth into households, we followed a two-stage randomization process. In the first stage all eligible households were stratified by community and randomized into treatment and control. We then conducted balance tests with the individual youth level data on several variables including sex, age, grade, working or not, in school or not, and a summary scores on each of the nine criteria. The randomization procedure was written to re-randomize until all variables balanced. In the case of Guyana, this resulted in 229 treatment and 228 control youth. This exceeds CFYR's capacity to treat, which was 150 youth. Therefore, the second step divides the treatment and control youth into primary study youth and replacement study youth. This step was done to maximize distribution of cases across communities and subsequently maximize power. In the case of Guyana, 150 youth were assigned to the primary study group

and the remaining 79 were assigned to the replacement group. Using Guyana as an example, Table 20 shows the number of youth by community randomized into treatment and control at both stage one and stage two.

Table 20: Results of stage 1 and 2 randomizations in Guyana

	Stage 1 randomization			Stage 2 randomization		
	Treatment	Control	Total	Treatment	Control	Total
Corriverton	43	43	86	30	30	60
East La Penitence	17	14	31	17	14	31
East Ruimveldt	41	41	82	30	33	63
Lodge	30	32	62	30	30	60
Sophia	98	98	196	43	43	86
Total	229	228	457	150	150	300

ANNEX II: IMPLEMENTATION DATES

Table 21 includes the implementation dates for the YSET and caregiver surveys, in each country and for each data collection period, and for the program. The *Family Matters* program officially started in June 2018 in Guyana, a month later in St. Lucia and the following month in St. Kitts and Nevis.

Table 21: YSET and caregiver survey implementation dates

	Program Implementation start date	YSET Survey Application Dates			Caregiver Survey Application Dates	
		Baseline	Midline	Endline	Baseline	Endline
Guyana	Jun-18	9/2017-3/2018	1/2019-2/2019	9/2019-10/2019	4/2018 - 6/2018	9/2019-10/2019
SLU	Jul-18	9/2017-4/2018	4/2019-6/2019	11/2019-12/2019	6/2018 – 7/2018	11/2019-12/2019
SKN	Aug-18	9/2017-6/2018	5/2019-6/2019	10/2019	7/2018 - 8/2018	11/2019

ANNEX III: BALANCE TESTS

YSET

Table 22 displays the balance at baseline between treatment and control youth in the ITT sample. The table compares the baseline means for the treatment and control groups across key variables. Non-zero differences between treatment and control means may imply systematic differences between treatment and control. The p-value is a number between 0 and 1 that denotes the likelihood of observing the difference between the treatment and control means if the true difference were zero. Therefore, higher p-values mean there is less evidence to reject the null hypothesis that the true difference is zero and lower p-values mean there is more evidence to reject the null hypothesis. Conservatively, we set a p-value threshold of 0.20. Any difference in means with a p-value above 0.20 means we fail to reject the null hypothesis that the true difference in means is zero, and any difference in means with a p-value below 0.20 means we reject the null hypothesis that the true difference in means is zero.

Column 1 reports the mean at baseline for all youth randomized to control. Column 2 reports the mean at baseline for all youth randomized to treatment, and column 3 reports the p-value from a t-test of means between control and treatment. The ITT sample is unbalanced for the mean scores for modules C and FSH. The sample is balanced in the key socio-demographic variables of age, male and out of school.

Table 22: Balance tests for treatment and control groups (ITT) at baseline, including false positives

Mean at Baseline (Std. Error)	Control (n=499)	Treatment (n=481)	P-Value
Age	13.81 (.093)	13.92 (.091)	0.390
Male	0.59 (.022)	0.60 (.022)	0.666
Out of school	0.10 (.013)	0.09 (.013)	0.734
Overall risk level (number of risk factors)	5.09 (.062)	5.20 (.064)	0.228
A: Anti-social tendencies (mean)	2.38 (.023)	2.37 (.024)	0.773
B: Weak parental monitoring (mean)	2.15 (.035)	2.12 (.036)	0.455
C: Critical life events (count)	3.71 (.061)	3.91 (.066)	0.028
DE: Impulsive risk taking (mean)	3.33 (.030)	3.37 (.031)	0.346
EG: Risky group behaviors (count)	2.10 (.087)	2.13 (.094)	0.798
F: Guilt neutralization (mean)	2.94 (.024)	2.92 (.022)	0.554
FSH: Horizontal family (mean)	2.35 (.028)	2.29 (.028)	0.187
FSV: Vertical family (mean)	2.57 (.038)	2.53 (.039)	0.368
G: Negative peer influence (mean)	1.95 (.040)	1.94 (.039)	0.888
H: Peer delinquency (mean)	1.90 (.028)	1.89 (.029)	0.963
IJ: Self-reported delinquency (count)	4.51 (.143)	4.51 (.149)	0.967
T: Family antisocial influence (sum)	5.89 (.115)	5.96 (.111)	0.641

Table 23 displays the balance at baseline between treatment and control youth in the TOT sample. Column 1 reports the mean at baseline for all youth randomized to control. Column 2 reports the mean at baseline for all youth randomized to treatment who took up and completed treatment. The TOT sample is unbalanced across the following categories: out of school, and the mean scores for modules C, DE, FSH and FSV.

Table 23: Balance tests for treatment and control groups (TOT) at baseline, including false positives

Mean at Baseline (Std. Error)	Control (n=499)	Treatment (n=230)	P-Value
Age	13.81 (.093)	13.77 (.126)	0.773
Male	0.59 (.022)	0.60 (.032)	0.706
Out of school	0.10 (.013)	0.06 (.016)	0.081
Overall risk level (number of risk factors)	5.09 (.062)	5.12 (.089)	0.790
A: Anti-social tendencies (mean)	2.38 (.023)	2.41 (.037)	0.515
B: Weak parental monitoring (mean)	2.15 (.035)	2.12 (.054)	0.597
C: Critical life events (count)	3.71 (.061)	3.87 (.095)	0.149
DE: Impulsive risk taking (mean)	3.33 (.030)	3.43 (.048)	0.076
EG: Risky group behaviors (count)	2.10 (.087)	2.03 (.129)	0.692
F: Guilt neutralization (mean)	2.94 (.024)	2.92 (.032)	0.646
FSH: Horizontal family (mean)	2.35 (.028)	2.27 (.042)	0.125
FSV: Vertical family (mean)	2.57 (.038)	2.48 (.058)	0.175
G: Negative peer influence (mean)	1.95 (.040)	1.87 (.054)	0.242
H: Peer delinquency (mean)	1.90 (.028)	1.86 (.042)	0.429
IJ: Self-reported delinquency (count)	4.51 (.143)	4.57 (.199)	0.810
T: Family antisocial influence (sum)	5.89 (.114)	5.90 (.154)	0.926

Table 24 displays the balance at baseline between the treatment and control youth in the matched TOT sample. Column 1 reports the mean at baseline for all youth randomized to control who were matched to treatment youth through our exact matching algorithm. Column 2 reports the mean at baseline for all youth randomized to treatment who took up and completed treatment. Taking again a p-value of 0.2 as a threshold for balance, the matched TOT sample is balanced across all categories. As expected, the means for some of the variables have greater differences between treatment and control in the matched versus the full TOT sample, but overall, there is greater balance in the matched sample because no variable is left unbalanced.

Table 24: Balance tests for matched treatment and control groups (TOT) at baseline, including false positives

Mean at Baseline (Std. Error)	Control (n=301)	Treatment (n=230)	P-Value
Age	13.85 (.118)	13.77 (.126)	0.638
Male	0.57 (.023)	0.60 (.032)	0.461
Out of school	0.09 (.016)	0.06 (.016)	0.218
Overall risk level (number of risk factors)	5.11 (.078)	5.12 (.089)	0.948
A: Anti-social tendencies (mean)	2.35 (.029)	2.41 (.037)	0.276
B: Weak parental monitoring (mean)	2.12 (.043)	2.12 (.054)	0.949
C: Critical life events (count)	3.75 (.078)	3.87 (.095)	0.328
DE: Impulsive risk taking (mean)	3.35 (.039)	3.43 (.048)	0.201
EG: Risky group behaviors (count)	2.00 (.104)	2.03 (.129)	0.816
F: Guilt neutralization (mean)	2.92 (.031)	2.92 (.032)	0.897
FSH: Horizontal family (mean)	2.33 (.034)	2.27 (.042)	0.295
FSV: Vertical family (mean)	2.57 (.050)	2.48 (.058)	0.255
G: Negative peer influence (mean)	1.96 (.052)	1.87 (.054)	0.241
H: Peer delinquency (mean)	1.88 (.035)	1.86 (.042)	0.710
IJ: Self-reported delinquency (count)	4.35 (.168)	4.57 (.199)	0.397
T: Family antisocial influence (sum)	6.04 (.147)	5.90 (.154)	0.530

CAREGIVER SURVEY

Table 25 shows balance at baseline across caregiver and youth characteristics for youth of caregivers in the ITT caregiver sample. Because the caregivers of some youth did not participate in the survey, the sample of families examined in the survey is slightly different than the YSET. As such, we revisit balance across control and treatment. As with the youth balance checks, we use a p-value of 0.20 as the cut off, meaning any characteristics with a p-value above 0.20 are balanced for treatment and control caregivers, and any characteristics with a p-value below 0.20 are unbalanced for treatment and control caregivers.

There are no statistically significant differences between treatment and control groups across the baseline characteristics of parental status, gender, age, education and employment status for caregivers and age, gender, overall risk, and modules A, B, DE, EG, F, FSV, G, H, IJ and T for youth. There are small, statistically significant differences between treatment and control caregivers on parental status. There are small, statistically significant differences between treatment and control youth on school status and modules C and FSH.

Table 25: Balance tests for treatment and control groups (ITT) at baseline, including false positives

Mean at Baseline (Std. Error)			
Caregiver characteristics	**Control (n=467)**	**Treatment (n=473)**	**P-Value**
Parental status	.850 (.017)	.821 (.019)	0.291
Male	.115 (.015)	.135 (.017)	0.360
Age	42 (.492)	43 (.510)	0.292
Education	3.67 (.038)	3.74 (.041)	0.218
Not employed	.632 (.023)	.635 (.024)	0.936
Youth characteristics	**Control (n=501)**	**Treatment (n=501)**	**P-Value**
Age	14 (.099)	14 (.098)	0.731
Male	.597 (.024)	.600 (.024)	0.935
Out of school	.104 (.015)	.094 (.014)	0.628
Overall risk level (number of risk factors)	5.14 (.067)	5.19 (.070)	0.579
A: Anti-social tendencies (mean)	2.37 (.025)	2.38 (.027)	0.841
B: Weak parental monitoring (mean)	2.16 (.038)	2.11 (.039)	0.353
C: Critical life events (count)	3.72 (.065)	3.90 (.071)	0.054
DE: Impulsive risk taking (mean)	3.33 (.032)	3.37 (.034)	0.301
EG: Risky group behaviors (count)	2.14 (.094)	2.14 (.102)	0.995
F: Guilt neutralization (mean)	2.93 (.026)	2.92 (.024)	0.719
FSH: Horizontal family (mean)	2.35 (.029)	2.27 (.030)	0.062
FSV: Vertical family (mean)	2.56 (.041)	2.52 (.042)	0.484
G: Negative peer influence (mean)	1.98 (.044)	1.91 (.042)	0.280
H: Peer delinquency (mean)	1.89 (.030)	1.90 (.032)	0.829
IJ: Self-reported delinquency (count)	4.55 (.156)	4.47 (.159)	0.694
T: Family antisocial influence (sum)	5.97 (.125)	5.93 (.120)	0.840

Table 26 shows balance at baseline across caregiver characteristics and across youth characteristics for youth of caregivers in the TOT caregiver sample. There are no statistically significant differences between treatment and control groups across baseline characteristics of gender, age, and employment status for caregivers and age, gender, overall risk, and most of the YSET modules. There are small, statistically significant differences between treatment and control caregivers on parental status. There are also small, statistically significant difference between treatment and control youth on school status and modules on critical life events (C), impulsive risk taking (DE), group of friends (FSH), and negative peer influence (G).

These differences are accounted for in the difference in difference estimation strategy. To further account for baseline differences in the TOT caregiver sample, we create a matched TOT sample using exact matching, following the same matching methodology used in the YSET analysis. As with the matched youth sample, we match the caregiver sample on youth characteristics.

Table 26: Balance tests for caregiver treatment and control groups (TOT) at baseline, including false positives

Mean at Baseline (Std. Error)			
Caregiver characteristics	**Control (n=467)**	**Treatment (n=224)**	**P-Value**
Parental status	.85 (.017)	.81 (.027)	0.239
Male	.115 (.015)	.117 (.022)	0.920
Age	42 (.492)	43 (.701)	0.270
Education	3.67 (.038)	3.81 (.063)	0.044
Not employed	.632 (.023)	.660 (.033)	0.4835
Youth characteristics	**Control (n=467)**	**Treatment (n=224)**	**P-Value**
Age	14 (.099)	14 (.131)	0.933
Male	.597 (.024)	.603 (.034)	0.883
Out of school	.104 (.015)	.065 (.017)	0.112
Overall risk level (number of risk factors)	5.14 (.066)	5.12 (.094)	0.869
A: Anti-social tendencies (mean)	2.37 (.025)	2.40 (.039)	0.518
B: Weak parental monitoring (mean)	2.16 (.038)	2.12 (.056)	0.597
C: Critical life events (count)	3.72 (.065)	3.90 (.100)	0.127
DE: Impulsive risk taking (mean)	3.33 (.032)	3.42 (.051)	0.107
EG: Risky group behaviors (count)	2.14 (.094)	2.02 (.135)	0.449
F: Guilt neutralization (mean)	2.93 (.026)	2.91 (.034)	0.627
FSH: Horizontal family (mean)	2.34 (.029)	2.24 (.042)	0.0418
FSV: Vertical family (mean)	2.56 (.041)	2.47 (.059)	0.231
G: Negative peer influence (mean)	1.98 (.044)	1.87 (.057)	0.127
H: Peer delinquency (mean)	1.89 (.030)	1.87 (.044)	0.660
IJ: Self-reported delinquency (count)	4.55 (.156)	4.60 (.209)	0.825
T: Family antisocial influence (sum)	5.97 (.125)	5.87 (.158)	0.639

ANNEX IV: OUTCOME VARIABLE CONSTRUCTION

YSET MODULE VARIABLES

The risk thresholds for each YSET module come from the USC determination of thresholds as explained in ANNEX I: ELIGIBILITY DETERMINATION AND RANDOMIZATION. These same thresholds for each module were used to generate module variables that show which youth is at risk for each module. Sometimes the thresholds are different for the different countries. The table below shows the YSET modules and their respective risk thresholds, and the YSET instrument is included in a supplementary annex. The scores are calculated by simply averaging the numbers associated with each answer for the modules where the mean method is used or counting the "yes" answers for the modules where the count method is used. If, for instance, a Guyanese youth answers "strongly agree" to all four questions in the module DE, their score for that module is (5+5+5+5)/4=5. Since 5 is above the threshold of 3 for module DE in Guyana, that youth would score at risk for that module.

Table 27: Risk thresholds and calculation method for YSET modules

		Risk Thresholds		
YSET module	**Method of calculation**	**Guyana**	**St. Lucia**	**St. Kitts and Nevis**
B	Mean	2.5	2	2.3
C	Count	4	4	4
DE	Mean	3	3	3.25
EG	Count	3	3	3
F	Mean	2.5	2.5	3
G	Mean	2.25	2.25	2
H	Mean	1.85	1.85	1.85
IJ	Count	3	3	3
T	Count	6	6	6

CAREGIVER SURVEY INDICATOR VARIABLES

Most caregiver indicators were generated using additive constructing and principal component analysis of the same variables. Caregiver outcomes of interest are:

1. Extended family cohesion index: Sum of Likert scale variables of how often youth visits any family members, calls any family members, and how often family members share family traditions. Variables are reverse coded, so higher values are positive.
2. Family Cohesion Scale: Factor of Likert scale variables of the extent to which the respondent agrees that family members are involved in each other's lives, family members feel very close, family members are supportive during difficult times, family members can calmly discuss problems

with each other, when angry, family members seldom say negative things about each other; and of Likert scale variables of the extent to which respondents are satisfied with the quality of communication between family members, family's ability to resolve conflicts and the amount of time spent together as a family. Variables are reverse coded, so higher values are positive.

3. Family consensus index: Sum of Likert scale variables of how often caregiver disagrees on how to raise or discipline youth; how often caregiver talks badly about other caregiver; and how often caregiver argues or fights in front of youth. Variables are reverse coded, so higher values are positive.
4. Parenting locus of control index: Sum of Likert scale variables of caregiver agreeing that children will get into trouble no matter what their parents do; the conditions in this community prevent me from keeping my children out of trouble; parents can have a major influence on their kids; and it is primarily my responsibility to keep my children out of trouble. The final two variables are reverse coded, so higher values are positive.
5. Improvement self-evaluation: Sum of Likert scale variables of caregiver agreeing that in the last 12 months, he/she has improved as a parent and in the past 12 months, his/her influence on his/her children has increased. Variables are reverse coded, so higher values are positive.
6. Caregiver presence index: Sum of continuous variables of the number of days per week the caregiver sits down to eat together with child; the number of days per week the caregiver helps child with schoolwork; the number of days per week the caregiver does something fun as a family with the child; and the number of days per week the caregiver does something religious as a family with the child.
7. Informed parent index: Average of Likert scale variables of how often child tells parent when he/she goes out, how often caregiver knows where child is when child is not at home or school, how often the caregiver knows who child is with when child is not at home or school, how often the child lets another adult know where he/she is if he/she does not let the caregiver know where he/she is. Variables are reverse coded, so high values are positive.
8. Parental authority index: Sum of Likert scale variables of how often there are consequences if family rules are broken, how often the child listens to the caregiver when the caregiver has something important to say, how often the caregiver listens to the child when the child has something important to say, and how often the child and caregiver are able to talk and solve their problems, how often the caregiver has other adults he/she can count on to help raise the child and how often the caregiver is proud of the child. The variables are reverse coded, so high values are positive.
9. Youth behavior and relationships index: Sum of when caregivers answer yes to whether they believe their child inhaled (sniffed, smoked or drank) drugs or substances that make you "high", skipped classes at school without an excuse or permission, lied about his/her age to get into some place or to buy something, purposely damaged or destroyed property that did not belong to him/her, carried a hidden weapon for protection, stolen or tried to steal something valuable, broken into a building to steal something, hit someone with the purpose of hurting him/her, attacked someone with a weapon, used a weapon or force to get money or things from people, been involved in group fights, been involved in gang fights, sold marijuana and illegal drugs.

For the Extended family cohesion index, Family adaptability and cohesion scale, Family consensus index, Parenting locus of control index, Improvement self-evaluation, Caregiver presence index, Informed parent index, and Parental authority index, higher values represent improved caregiver outcomes. For the Youth behavior index, lower values represent improved outcomes.

COMMUNITY SURVEY INDICATOR VARIABLES

Principal component analysis (PCA) is a data reduction method: from a group of correlated variables; it produces a smaller set of components that give a good representation of the main characteristics of the original data. These components are a linear combination of the original variables that represent the maximum possible variance between them but are uncorrelated to each other (whereas there may be a high degree of correlation among the original variables).

Described mathematically, given an original set of *n* correlated variables, PCA produces uncorrelated, orthogonal components, where each component is a linear combination of the initial variables. For example, from a set of variables X_1 through to X_n,

$$1^{st} \text{ Principal Component} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1n}X_n$$

$$2^{nd} \text{ Principal Component} = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2n}X_n$$

...

$$Z^{th} \text{ Principal Component} = a_{Z1}X_1 + a_{Z2}X_2 + \ldots + a_{Zn}X_n$$

where a_{Zn} is the coefficient, or weight, for the Zth principal component and the nth variable. Each component has both magnitude and direction

The first component accounts for maximum possible variance in the data. Subsequent components account for information not captured by the first and are uncorrelated with the first component (and each other), with diminishing explanatory power of variance among the original variables. Therefore, the first component often can reasonably be used on its own as an index score that gives information on the underlying variables.

PCA does not assume the existence of a group of common factors driving the variation in the data (whereas factor analysis does). It does, however, assume that the underlying data is normally distributed and calculates associations between variables using Pearson correlations. This means that PCA on its own is best suited for continuous data. Conversely, the data from which the indices in the **Error! Reference source not found.** section of this report are built are ordinal, categorical data.

PCA with polychoric correlations solves this problem by assuming that each ordinal variable is a representation of an underlying, inferred "latent variable" that is both continuous and normally distributed, and produces principal components under this assumption. When polychoric correlations are used in variables with many categories, the method assumes that a discrete amount is actually a continuous variable. In the case of the Likert scales or categorical dichotomic variables, the assumption is more likely to follow when the realizations are not particularly skewed or lopsided in their distribution.

Table 28 shows the variables from the community survey that form each index in the community context section.

Table 28: Questions used for each community survey index

Index Name	1st Principal Component	Question numbers
Perception of Crime Prevalence	crimeprev	B20
		B21
		B22
		B23
		B24 (recode)
		B25
		B26
Social Capital I	soccap1	B1
		B2 (recode)
		B4
Social Capital II	soccap2	B36
		B37
		B38
		B39
		B40
		B41
Attitudes Towards the Law	lawatt	B5 (recode)
		B6 (recode)
		B7
		B8 (recode)
		B9 (recode)
		B10 (recode)
View of The Police	police	B11
		B13
		B14
		B27 (recode)
Resources for Youth	resources	B12
		B15
		B16
		B18
		B19
Self-Efficacy	selfeff	D1 (recode)
		D2 (recode)
		D3 (recode)
		D4 (recode)
		D5 (recode)
		D6
		D7

ANNEX V: ADDITIONAL YSET SURVEY RESULTS

ANALYSIS RESULTS: ITT, TOT FULL SAMPLE, AND TOT MATCHED SAMPLE

Ideally, the ITT sample would include data on all youth that were randomized, but because of survey attrition, the sample usually becomes smaller. In this case, 167 youth dropped out of the survey, which amounts to 15 percent of the sample (see Figure 3 in the "Sample description" section for a clear picture of survey attrition).

Table 29 and Table 30 show the coefficients for the simple difference-in-differences regression model, and the preferred model with random effects, clustered standard errors and community dummies. The "mean number of risk factors," "four or more risk factors," scale T outcome variables, and the average scores for the DE and T scales have statistically significant results in this sample.

Table 29: Analysis results on key YSET outcomes for ITT sample (n=980)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
Mean Number of Risk Factors	-0.345* (0.18)	-0.345** (0.15)
Four Or More Risk Factors	-0.062* (0.04)	-0.062* (0.03)
B: Weak Parental Monitoring	-0.052 (0.04)	-0.052 (0.04)
C: Critical Life Events	-0.037 (0.04)	-0.037 (0.04)
DE: Impulsive Risk Taking	-0.042 (0.04)	-0.042 (0.04)
EG: Risky Behaviors	-0.022 (0.04)	-0.022 (0.04)
F: Guilt Neutralization	-0.053 (0.04)	-0.053 (0.04)
G: Negative Peer Influence	-0.024 (0.04)	-0.024 (0.04)
H: Peer Delinquency	-0.004 (0.04)	-0.004 (0.04)
IJ: Self-Reported Delinquency	-0.016 (0.04)	-0.016 (0.04)
T: Family Antisocial Influence	-0.095** (0.04)	-0.095** (0.04)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 30: Analysis results on key YSET outcomes for ITT sample (n=980)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
A: Anti-Social Tendencies (Mean)	-0.018 (0.05)	-0.018 (0.04)
B: Weak Parental Monitoring (Mean)	-0.075 (0.07)	-0.074 (0.06)
C: Critical Life Events (Count)	-0.193 (0.14)	-0.193 (0.12)
DE: Impulsive Risk Taking (Mean)	-0.115 (0.07)	-0.115* (0.06)
EG: Risky Behaviors (Count)	-0.153 (0.18)	-0.153 (0.15)
F: Guilt Neutralization (Mean)	-0.053 (0.05)	-0.053 (0.05)
FSH: Horizontal Family (Mean)	0.006 (0.06)	0.006 (0.05)
FSV: Vertical Family (Mean)	0.054 (0.08)	0.054 (0.06)
G: Negative Peer Influence (Mean)	-0.077 (0.08)	-0.077 (0.07)
H: Peer Delinquency (Mean)	0.013 (0.06)	0.013 (0.05)
IJ: Self-Reported Delinquency (Count)	-0.233 (0.31)	-0.233 (0.25)
T: Family Antisocial Influence (Sum)	-0.406* (0.24)	-0.406** (0.20)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 31 and Table 32 show the coefficients for the regressions with the full TOT sample. The results are similar for the two models, with the preferred model having more statistically significant results for "mean number of risk factors," scale T, and scale DE.

Table 31: Analysis results on key YSET outcomes for TOT sample (n=729)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
Mean Number of Risk Factors	-0.418* (0.22)	-0.418** (0.19)
Four Or More Risk Factors	-0.093** (0.04)	-0.093** (0.04)
B: Weak Parental Monitoring	-0.070 (0.06)	-0.070 (0.05)
C: Critical Life Events	0.005 (0.06)	0.005 (0.05)
DE: Impulsive Risk Taking	-0.069 (0.05)	-0.069 (0.05)
EG: Risky Behaviors	-0.041 (0.05)	-0.041 (0.04)
F: Guilt Neutralization	-0.090* (0.05)	-0.090* (0.05)
G: Negative Peer Influence	-0.025 (0.05)	-0.025 (0.05)
H: Peer Delinquency	0.007 (0.05)	0.007 (0.05)
IJ: Self-Reported Delinquency	-0.037 (0.05)	-0.037 (0.05)
T: Family Antisocial Influence	-0.098* (0.06)	-0.098** (0.05)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 32: Analysis results on key YSET outcomes for TOT sample (n=729)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
A: Anti-Social Tendencies (Mean)	-0.020 (0.06)	-0.020 (0.05)
B: Weak Parental Monitoring (Mean)	-0.121 (0.09)	-0.121 (0.08)
C: Critical Life Events (Count)	-0.093 (0.17)	-0.093 (0.16)
DE: Impulsive Risk Taking (Mean)	-0.160* (0.09)	-0.160** (0.08)
EG: Risky Behaviors (Count)	-0.166 (0.22)	-0.166 (0.17)
F: Guilt Neutralization (Mean)	-0.112* (0.07)	-0.112* (0.06)
FSH: Horizontal Family (Mean)	-0.011 (0.07)	-0.011 (0.06)
FSV: Vertical Family (Mean)	0.087 (0.10)	0.089 (0.08)
G: Negative Peer Influence (Mean)	-0.074 (0.10)	-0.074 (0.08)
H: Peer Delinquency (Mean)	0.002 (0.07)	0.002 (0.06)
IJ: Self-Reported Delinquency (Count)	-0.368 (0.38)	-0.368 (0.31)
T: Family Antisocial Influence (Sum)	-0.254 (0.30)	-0.254 (0.24)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 33 and Table 34 show the coefficients for the regressions with the matched TOT sample. The results are very similar for the two models, with the preferred model having more statistically significant results for the mean scores on scales B, DE, and F.

Table 33: Analysis results on key YSET outcomes for matched TOT sample (n=531)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
Mean Number of Risk Factors	-0.314 (0.24)	-0.314 (0.21)
Four Or More Risk Factors	-0.052 (0.05)	-0.052 (0.05)
B: Weak Parental Monitoring	-0.065 (0.06)	-0.065 (0.05)
C: Critical Life Events	0.029 (0.06)	0.029 (0.06)
DE: Impulsive Risk Taking	-0.062 (0.06)	-0.062 (0.05)
EG: Risky Behaviors	-0.048 (0.05)	-0.048 (0.05)
F: Guilt Neutralization	-0.101* (0.06)	-0.101* (0.06)
G: Negative Peer Influence	-0.011 (0.06)	-0.011 (0.05)
H: Peer Delinquency	0.003 (0.06)	0.003 (0.05)
IJ: Self-Reported Delinquency	-0.023 (0.06)	-0.023 (0.05)
T: Family Antisocial Influence	-0.035 (0.06)	-0.035 (0.05)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 34: Analysis results on key YSET outcomes for matched TOT sample (n=531)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
A: Anti-Social Tendencies (Mean)	-0.034 (0.07)	-0.034 (0.06)
B: Weak Parental Monitoring (Mean)	-0.157 (0.10)	-0.157* (0.08)
C: Critical Life Events (Count)	-0.017 (0.19)	-0.017 (0.17)
DE: Impulsive Risk Taking (Mean)	-0.189* (0.10)	-0.189** (0.08)
EG: Risky Behaviors (Count)	-0.205 (0.23)	-0.205 (0.19)
F: Guilt Neutralization (Mean)	-0.132* (0.07)	-0.132** (0.07)
FSH: Horizontal Family (Mean)	-0.034 (0.08)	-0.034 (0.07)
FSV: Vertical Family (Mean)	0.021 (0.11)	0.023 (0.09)
G: Negative Peer Influence (Mean)	-0.070 (0.11)	-0.070 (0.09)
H: Peer Delinquency (Mean)	0.007 (0.08)	0.007 (0.06)
IJ: Self-Reported Delinquency (Count)	-0.346 (0.40)	-0.346 (0.33)
T: Family Antisocial Influence (Sum)	-0.083 (0.33)	-0.083 (0.27)

Note: *** p<0.01, ** p<0.05, * p<0.1

The primary effects of the *Family Matters* program on risk outcomes are displayed in Table 35 and Table 36. The analysis compares eligible youth assigned to treatment and eligible youth assigned to control in each of the three samples, the ITT, the TOT full sample, and the TOT matched sample. We report coefficients and standard errors for our preferred model, with community dummies, random effects and clustered standard errors at the individual level.[13] The results show that program effects are strongest in the full TOT sample, where the coefficients for "mean number of risk factors" and "four or more risk factors" are larger and more statistically significant than in the other two samples.

Table 35: Sample comparison of results using difference-in-differences model with random effects and clustered standard errors

	ITT	TOT Full Sample	TOT Matched Sample
n	C=499 T=481	C=499 T=230	C=301 T=230
Mean Number of Risk Factors	**-0.345**** **(0.15)**	**-0.418**** **(0.19)**	-0.314 (0.21)
Four or More Risk Factors	**-0.062*** **(0.03)**	**-0.093**** **(0.04)**	-0.052 (0.05)
B: Weak Parental Monitoring	-0.052 (0.04)	-0.070 (0.05)	-0.065 (0.05)
C: Critical Life Events	-0.037 (0.04)	0.005 (0.05)	0.029 (0.06)
DE: Impulsive Risk Taking	-0.042 (0.04)	-0.069 (0.05)	-0.062 (0.05)
EG: Risky Behaviors	-0.022 (0.04)	-0.041 (0.04)	-0.048 (0.05)
F: Guilt Neutralization	-0.053 (0.04)	**-0.090*** **(0.05)**	**-0.101*** **(0.06)**
G: Negative Peer Influence	-0.024 (0.04)	-0.025 (0.05)	-0.011 (0.05)
H: Peer Delinquency	-0.004 (0.04)	0.007 (0.05)	0.003 (0.05)
IJ: Self-Reported Delinquency	-0.016 (0.04)	-0.037 (0.05)	-0.023 (0.05)
T: Family Antisocial Influence	**-0.095**** **(0.04)**	**-0.098**** **(0.05)**	-0.035 (0.05)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 36: Sample comparison of results using difference-in-differences model with random effects and clustered standard errors

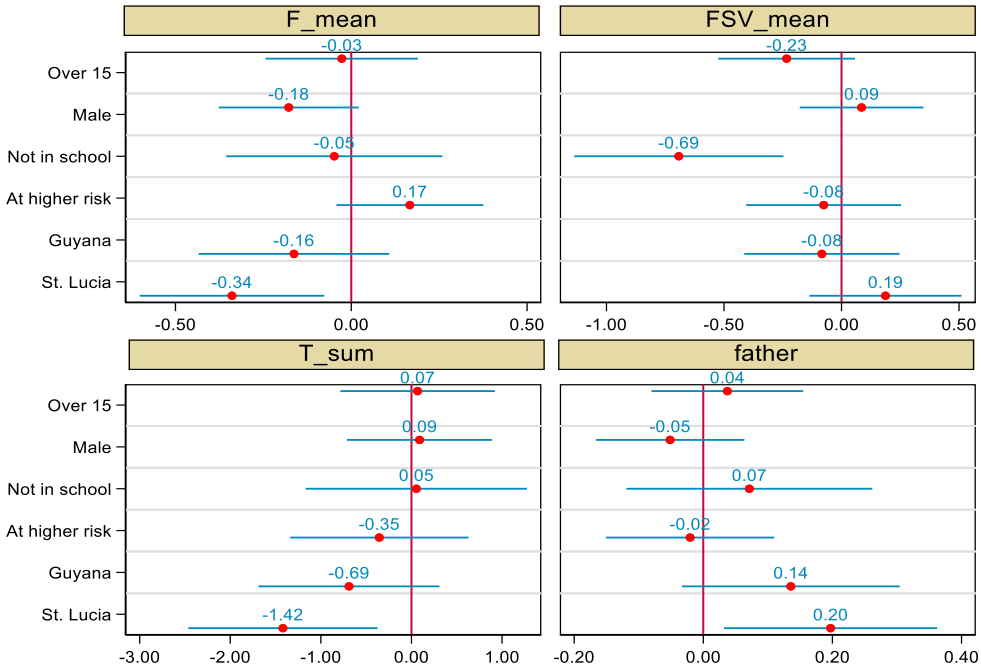
	ITT	TOT Full Sample	TOT Matched Sample
n	C=499 T=481	C=499 T=230	C=301 T=230
A: Anti-Social Tendencies (Mean)	-0.018 (0.04)	-0.020 (0.05)	-0.034 (0.06)
B: Weak Parental Monitoring (Mean)	-0.074 (0.06)	-0.121 (0.08)	**-0.157*** **(0.08)**
C: Critical Life Events (Count)	-0.193 (0.12)	-0.093 (0.16)	-0.017 (0.17)
De: Impulsive Risk Taking (Mean)	**-0.115*** **(0.06)**	**-0.160**** **(0.08)**	**-0.189**** **(0.08)**
EG: Risky Behaviors (Count)	-0.153 (0.15)	-0.166 (0.17)	-0.205 (0.19)
F: Guilt Neutralization (Mean)	-0.053 (0.05)	**-0.112*** **(0.06)**	**-0.132**** **(0.07)**
FSH: Horizontal Family (Mean)	0.006 (0.05)	-0.011 (0.06)	-0.034 (0.07)
FSV: Vertical Family (Mean)	0.054 (0.06)	0.089 (0.08)	0.023 (0.09)
G: Negative Peer Influence (Mean)	-0.077 (0.07)	-0.074 (0.08)	-0.070 (0.09)
H: Peer Delinquency (Mean)	0.013 (0.05)	0.002 (0.06)	0.007 (0.06)
IJ: Self-Reported Delinquency (Count)	-0.233 (0.25)	-0.368 (0.31)	-0.346 (0.33)
T: Family Antisocial Influence (Sum)	**-0.406**** **(0.20)**	-0.254 (0.24)	-0.083 (0.27)

Note: *** p<0.01, ** p<0.05, * p<0.1

HETEROGENEOUS EFFECTS

Figure 46 shows additional heterogeneous treatment effects results. For the mean of the F scale, the risk reduction for youth in St. Lucia is statistically different from youth in St. Kitts and Nevis (the base country for these charts), at the 90 percent confidence level. The average difference-in-difference score for youth in St Lucia is smaller by 0.34 points, compared to St. Kitts and Nevis. In the FSV scale, the mean score is smaller for youth out of school, compared to youth who were in school. In the T scale, the result for St. Lucia is statistically different from St. Kitts and Nevis. For the father outcome variable, youth in St. Lucia were 20 percent more likely to have a father at endline compared with youth in St. Kitts and Nevis (in terms of their difference-in-difference scores, which subtracts the control changes from baseline to endline).

Figure 46: Heterogeneous treatment effects for the F, FSV and T scales, and the father outcome variable



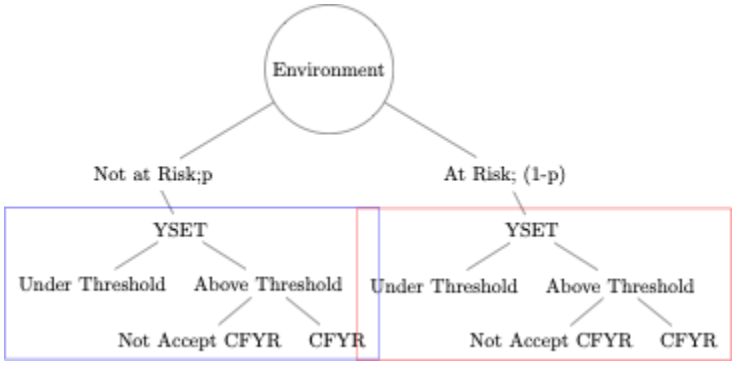
ANNEX VI: THEORETICAL FRAMEWORK REGARDING OBSERVABILITY AND IDENTIFYING CAUSAL EFFECTS

The ways in which the CFYR counseling program can change behavior can be depicted visually through a decision tree that highlights the many possible decision paths that may influence the observable data collected in this evaluation. Several of the paths in the decision tree are relevant for the overall dynamics of youth behavior in their families, schools and neighborhoods, but they cannot be directly observed (or at least not within the research setup designed for this study). Knowing that there are some behaviors that are not directly observable does not undermine the validity of the findings in the evaluation. If anything, to the extent that there is a good understanding of the framework set up for the evaluation, where random assignment created a treatment and a control group of youth at risk, a better understanding of the context in which youth behave, which includes many aspects that are outside of the evaluation, only strengthens the confidence in our findings.

OBSERVABLE DECISIONS AND RISK

If we start from a simple dichotomic classification of the situation of youth, we can think of them as being assigned by a "state of nature" into two possible situations, being at risk or not at risk. Clearly, youth themselves can play an important role in how likely they are to fall into each of these two states, but by considering these to be a probabilistic function of a given "natural" environment, one can simplify the complex societal, cultural, economic and institutional features that may impact the youth's situation. Let us further assume that it is possible to assign a probability *p* to the situation of a youth not being at risk, and a complementary probability *(1-p)* of being at risk. One can imagine the CFYR program as a tree that starts from this initial condition, like the one depicted in Figure 47.

Figure 47: Simplified CFYR Program Tree

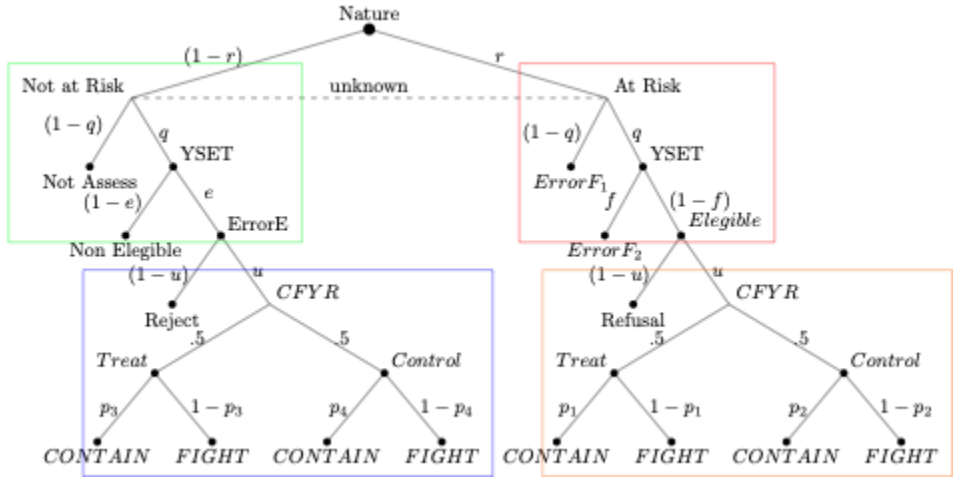


In this initial diagram, there are two possible groups of youth to whom the YSET is applied as an instrument to measure risk. The ones on the left colored in blue, in principle, should not be included in the intervention, while the ones on the right are the at-risk youth that should be identified as eligible by the instrument. The instrument should find that the youth on the left will be under the threshold, and therefore not be offered to become part of CFYR. The youth on the right will be above the threshold and

offered to be part of the program. There is still a final branch where eligible youth and their families would decide whether or not they accept to be in the program. In principle, if there are no targeting errors, most youth who are not at risk will be under the threshold. If the YSET is a good tool to diagnose risk, most youth on the right side would be offered the intervention, and if they and their families can see the benefits that the program offers, would accept to be part of it.

This simplified version of the decision tree does not, however, really depict the way the program has worked on the ground. A more realistic set of choices and decisions can be illustrated by Figure 48.

Figure 48: Realistic CFYR Program Tree



The number of youth that are incorrectly enrolled into the program (Error E) are:

$$[(1-r)qe]N$$

Some of them might reject the program, seeing little benefit in something they do not deem they need. For simplicity we will assume that the uptake *(u)* of the program is the same for both youth at risk and not at risk. Therefore, the probability of declining participation is given by *1-u*. One can interpret this equal uptake on both branches of the tree as a consequence of offsetting incentives that may make both groups equally likely to participate. In principle, youth who are not at risk would reject the intervention and those at risk would accept it. But if the counseling services are valuable, youth who are not at risk may want to take advantage of all the opportunities and services available in their community, even when they do not really need them. And youth and families at risk may be reluctant to accept help when they in fact need it, due to psychological mechanisms of denial.

Therefore, the number of youth not at risk that receive the program is:

$$[(1-r)qeu]N$$

This group is critical, particularly if it is large, because it may influence the way in which the average treatment effects of the program actually play out.

There is a different type of mistake in targeting that is made up of the youth that are at risk, but do not get identified to be surveyed with the YSET (perhaps because they are not enrolled at school, and they cannot even be easily found in their neighborhoods because they may be living in more than one home

with little caregiving) and those that are incorrectly classified as not risky by the YSET. Those two errors (F1 and F2) together make up the following group:

$$[r(1-q)+rqf]N = rN[1-q(1+f)]$$

Which simply tells us that the error of Failure to target (Error F) is the share of at-risk youth that did not get tested by the YSET or scored under the risk threshold of the instrument.

The program incorporates youth voluntarily, so there is a share *(u)* that takes up the intervention, and (*1-u)* that refuses. We assume that this share is the same for both at-risk and not-at-risk youth as a simplification, which yields the following at-risk youth in the program:

$$[rq(1-f)u]N$$

Hence the total number of both at-risk and not-at-risk youth in the CFYR program are:

$$\{[(1-r)qeu]+[rq(1-f)u]\}N = [e+r-r(f+e)]uqN$$

This tells us that the coverage of the program can be decomposed in two parts. It depends on *uq*, which represents the share of youth that are successfully contacted to take the YSET, multiplied by the share that have accepted uptake, who become part of CFYR. The second part includes the youth at risk, the ones that have been erroneously given the intervention without requiring it, corrected by the two types of mistakes *(f and* e*)* in the program.

The final branches of the tree show the randomization done for the evaluation, simplified to assume that exactly half of the youth are assigned with a probability 0.5 to treatment and the other half to control. Once the counseling ends, each of these groups of youth have probabilities *p1-p4* of containing antisocial behavior. To simplify, we call the risky behavior "fight" and the more positive behavior "contain." If the intervention is successful, *p1>p2*. That is, a youth at risk with intervention is more likely to contain antisocial behavior than the control group.

These expressions may be useful to explain observed differences in the frequency of changes in antisocial behavior in the evaluation. The simple way to think about this is that there should be a difference in the size of the groups that contain their antisocial behavior between treatment and control, namely:

$$0.5\{p_3[(1-r)qeu]+p_1[rq(1-f)u]\} > 0.5\{p_4[(1-r)qeu]+p_2[rq(1-f)u]\}$$

Simplifying for *.5qu* which is the same across the board, we get:

$$p_3[(1-r)e]+p_1[r(1-f)] > p_4[(1-r)e]+p_2[r(1-f)]$$

Regrouping:

$$(p_1-p_2) > (p_4-p_3)\frac{(1-r)}{r}\frac{e}{(1-f)}$$

Hence, we can know that the reasons why we may not observe a difference between the treated and the control group may stem from three sources, assuming that there is an effect of the intervention on the youth at risk: 1) it may be that the effect of the treated youth not at risk is counterproductive (i.e. p4 being larger than p3), 2)the differences between *p* may be very small, and 3) the right-hand side gets larger

than one as the odds ratio of youth not being at risk is higher, namely in communities with higher probabilities of finding that youth are not at risk. The ratio of the probability of incorporating in the program false positives compared to the probability of having truly at-risk youth in the program increases, it is more likely that no differences may be observed (i.e. the more the program fails to incorporate youth compared to how much inclusion of false positives it has).

Our study can measure some cases that are demonstrably "false positives." These would be cases when youth are included in the program by mistake even though the YSET determined they should not be in the program. However, that is not the full extent of potential false positives, since it is possible the YSET itself, as an instrument, makes errors of categorization of risk.[22]

CAUSAL GRAPHS TO ILLUSTRATE THE IDENTIFICATION CHALLENGE

The identification challenge is to estimate the effect of CFYR on behavior that may involve antisocial or criminal activity and gang membership. Simply regressing antisocial behavior on CFYR may not provide an unbiased estimate of the causal effect of the program, because selection into CFYR was driven by risk, as measured and identified by the YSET instrument, as well as self-selection into the program. Youth that accept the intervention may be more likely, for example, to be the ones that would have presented the attitudinal characteristics and grit to prevent them, absent the CFYR intervention, from engaging in undesirable behavior to begin with.

It is possible to express this as a structural equation model, which can clarify some of the difficulties in causal identification. We can have $RISK_0$ representing the true (unobserved) risk prior to the intervention and $RISK_1$ the true (also unobserved) risk afterwards. CFYR includes attributes of the counseling intervention that is geared towards reducing that risk, and B is the behavior that the program seeks to change.

1. $RISK_0 = U_1$
2. $CFYR = a\ RISK_0 + U_2$
3. $RISK_1 = c\ RISK_0 + d\ CFYR + U_3$
4. $BEHAVIOR = b\ CFYR + e\ RISK_1 + U_4$

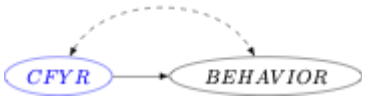
Such a structural equation model would suggest that it is not possible to identify correctly the effect of CFYR on B, unless something else is known about how to model the effects of the latent variable of risk. In other words, claiming that on average youth who participated in CFYR did not engage in risky behavior (which is what a coefficient of a simple regression of those variables would yield), does not provide a sound causal inference. Running the following model:

$CFYR = U_C$

$BEHAVIOR = b\ CFYR + U_B$

[22] We have 32 of them in the panel TOT sample (our main sample of analysis), with 23 in the control group and nine in the treatment group. Results do not really change when we exclude the false positives, except for a few decreases in statistical significance, so we are running all of the analysis, including the false positives.

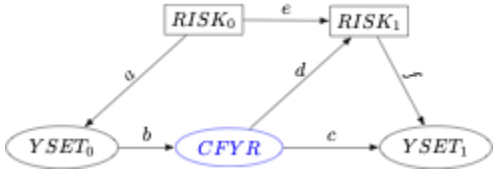
Figure 49: Simple Model of Behavior



CFYR's effect on behavior (Figure 49) is not identified, because C would be endogenous, and the error U_B would be correlated with the error of U_C. The actual model we estimate, in which we remove the latent variables of risk, is actually something like the following structural equation (Figure 50):

1. $RISK_0 = U_1$
2. $YSET_0 = a\ RISK_0 + U_2$
3. $CFYR = b\ YSET_0 + U_3$
4. $RISK_1 = e\ RISK_0 + d\ CFYR + U_4$
5. $YSET_1 = c\ CFYR + f\ RISK_1 + U_5$

Figure 50: Structural Equation Model



Removing the latent variables (1) and (4) yields:

1. $YSET_0 = a\ RISK_0 + U_Y$
2. $CFYR = b\ YSET_0 + U_C$
3. $YSET_1 = b\ CFYR + U_S$

where b = c + df, which means that regressing YSET at endline on YSET at baseline and the CFYR intervention yields a coefficient of a regression which tells us about both the direct effect of the intervention as well as the indirect net effect of the underlying true (unobservable) risk.

In the end, this evaluation was based on an RCT design, in which identification is ensured from the random assignment of youth into CFYR treatment and control groups. This ensures that, provided there is balance, we are measuring a LATE of the counseling intervention.

ANNEX VII: ADDITIONAL CAREGIVER SURVEY RESULTS

MAIN EFFECTS

Table 37: Analysis results on key caregiver outcomes for ITT caregiver sample (n=919)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
Extended family cohesion index	-0.168 (0.29)	-0.168 (0.24)
Family cohesion scale	-0.007 (0.09)	-0.012 (0.07)
Family consensus index	0.263 (0.28)	0.263 (0.25)
Parenting locus of control index	-0.100 (0.18)	-0.100 (0.16)
Improvement self-evaluation	0.019 (0.14)	0.019 (0.12)
Caregiver presence index	-0.772 (0.62)	-0.772 (0.49)
Informed parent index	-0.103 (0.10)	-0.103 (0.07)
Parental authority index	-0.180 (0.34)	-0.180 (0.30)
Youth behavior and relationship index	0.031 (0.11)	0.031 (0.09)

Table 38: Analysis results on key caregiver outcomes for TOT caregiver sample (n=695)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
Extended family cohesion index	-0.073 (0.36)	-0.073 (0.29)
Family cohesion scale	-0.051 (0.11)	-0.060 (0.09)
Family consensus index	-0.091 (0.35)	-0.091 (0.32)
Parenting locus of control index	-0.096 (0.22)	-0.096 (0.20)
Improvement self-evaluation	0.014 (0.16)	0.014 (0.13)
Caregiver presence index	-0.425 (0.76)	-0.425 (0.59)
Informed parent index	-0.145 (0.12)	-0.142 (0.09)
Parental authority index	-0.331 (0.42)	-0.331 (0.38)
Youth behavior and relationship index	0.013 -0.073	0.013 (0.12)

Table 39: Analysis results on key caregiver outcomes for matched TOT caregiver sample (n=518)

	Simple Diff-In-Diff	With Random Effects and Clustered Standard Errors
Extended family cohesion index	0.017 (0.39)	0.017 (0.32)
Family cohesion scale	0.069 (0.12)	0.059 (0.10)
Family consensus index	-0.203 (0.39)	-0.203 (0.34)
Parenting locus of control index	-0.104 (0.24)	-0.092 (0.21)
Improvement self-evaluation	-0.021 (0.17)	-0.004 (0.15)
Caregiver presence index	-0.098 (0.86)	-0.098 (0.66)
Informed parent index	-0.148 (0.13)	-0.144 (0.10)
Parental authority index	-0.327 (0.46)	-0.327 (0.42)
Youth behavior and relationship index	0.007 (0.16)	0.007 (0.13)

Table 40: Sample comparison of results using difference in difference model with random effects

	ITT Sample (n=919)	TOT Sample (n=695)	Matched TOT Sample (n=518)
Extended family cohesion index	-0.168 (0.24)	-0.073 (0.29)	0.017 (0.32)
Family cohesion scale	-0.012 (0.07)	-0.060 (0.09)	0.059 (0.10)
Family consensus index	0.263 (0.25)	-0.091 (0.32)	-0.203 (0.34)
Parenting locus of control index	-0.100 (0.16)	-0.096 (0.20)	-0.092 (0.21)
Improvement self-evaluation	0.019 (0.12)	0.014 (0.13)	-0.004 (0.15)
Caregiver presence index	-0.772 (0.49)	-0.425 (0.59)	-0.098 (0.66)
Informed parent index	-0.103 (0.07)	-0.142 (0.09)	-0.144 (0.10)
Parental authority index	-0.180 (0.30)	-0.331 (0.38)	-0.327 (0.42)
Youth behavior and relationship index	0.031 (0.09)	0.013 (0.12)	0.007 (0.13)

HETEROGENEOUS EFFECTS

Figure 51 and Figure 52 present differential program effects on caregiver indicators including extended family cohesion index, family cohesion index, parental authority index, youth behavior and relationships index parenting locus of control index and improvement self-evaluation index by varied groups, decomposed by these factors. We do not see differential impacts across most indicators.

Figure 51: Heterogeneous treatment effects on Extended Family Cohesion Index, Family Cohesion Index, Parental Authority Index and Youth Behavior and Relationships Index

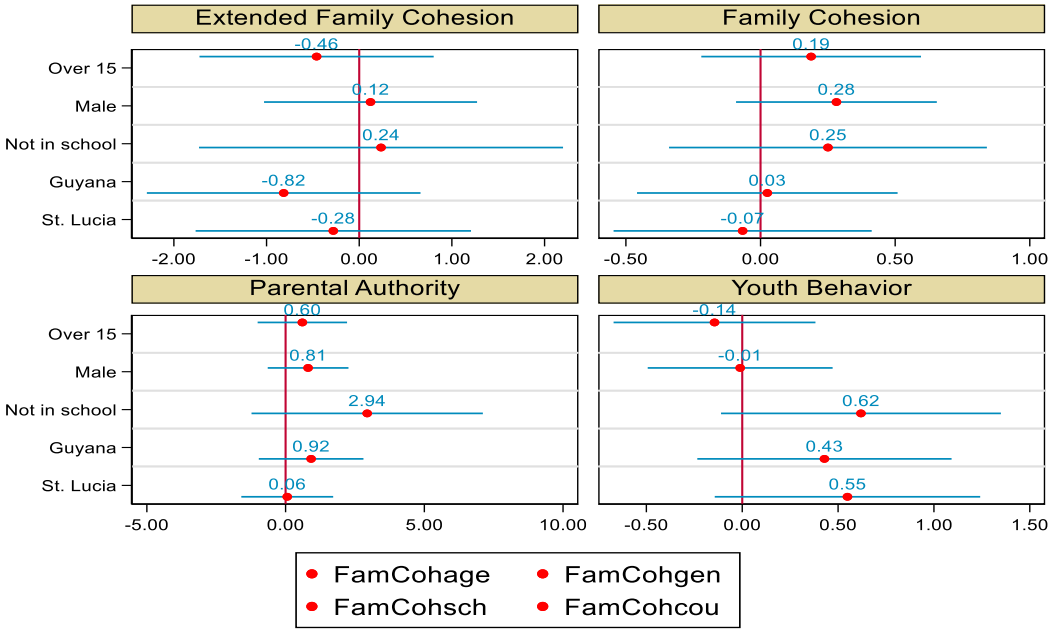
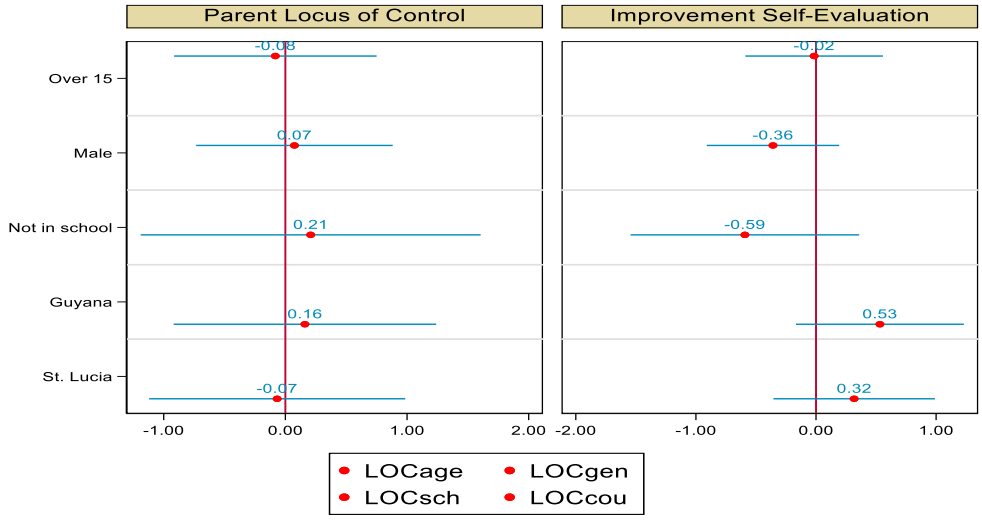


Figure 52: Heterogeneous treatment effects on Parenting Locus of Control Index and Improvement Self-Evaluation



ANNEX VIII: ADDITIONAL CHARTS

YSET

Figure 53: Youth responses on questions from Module B (n=729)

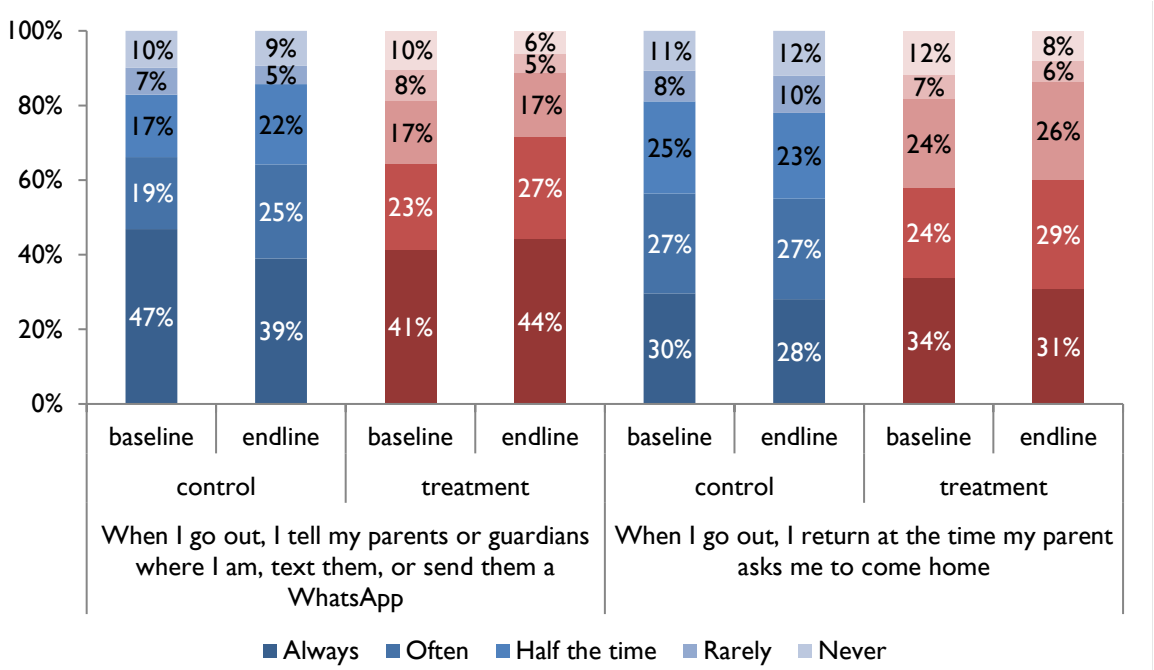


Figure 54: Youth responses on questions from Module B (n=729)

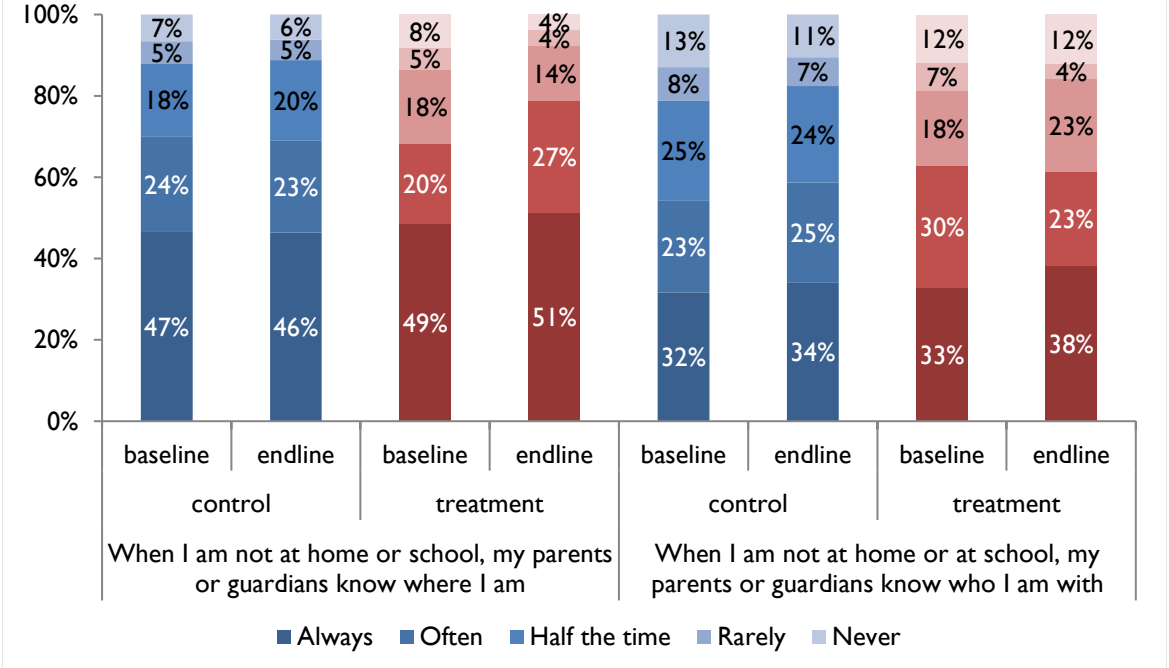


Figure 55: Youth responses on questions from Module B (n=729)

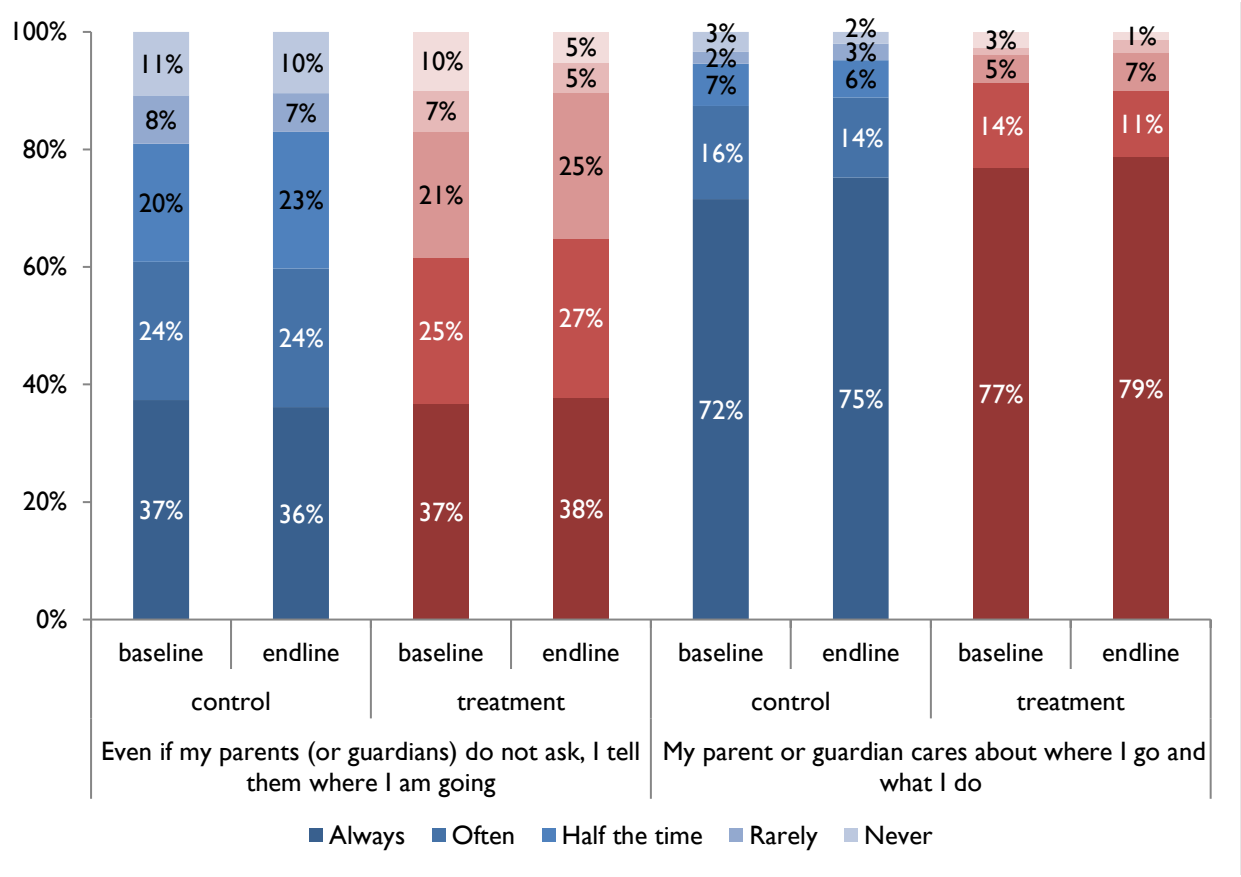


Figure 56: Share of youth responding yes to behavioral questions, more commonly reported behaviors (n=729)

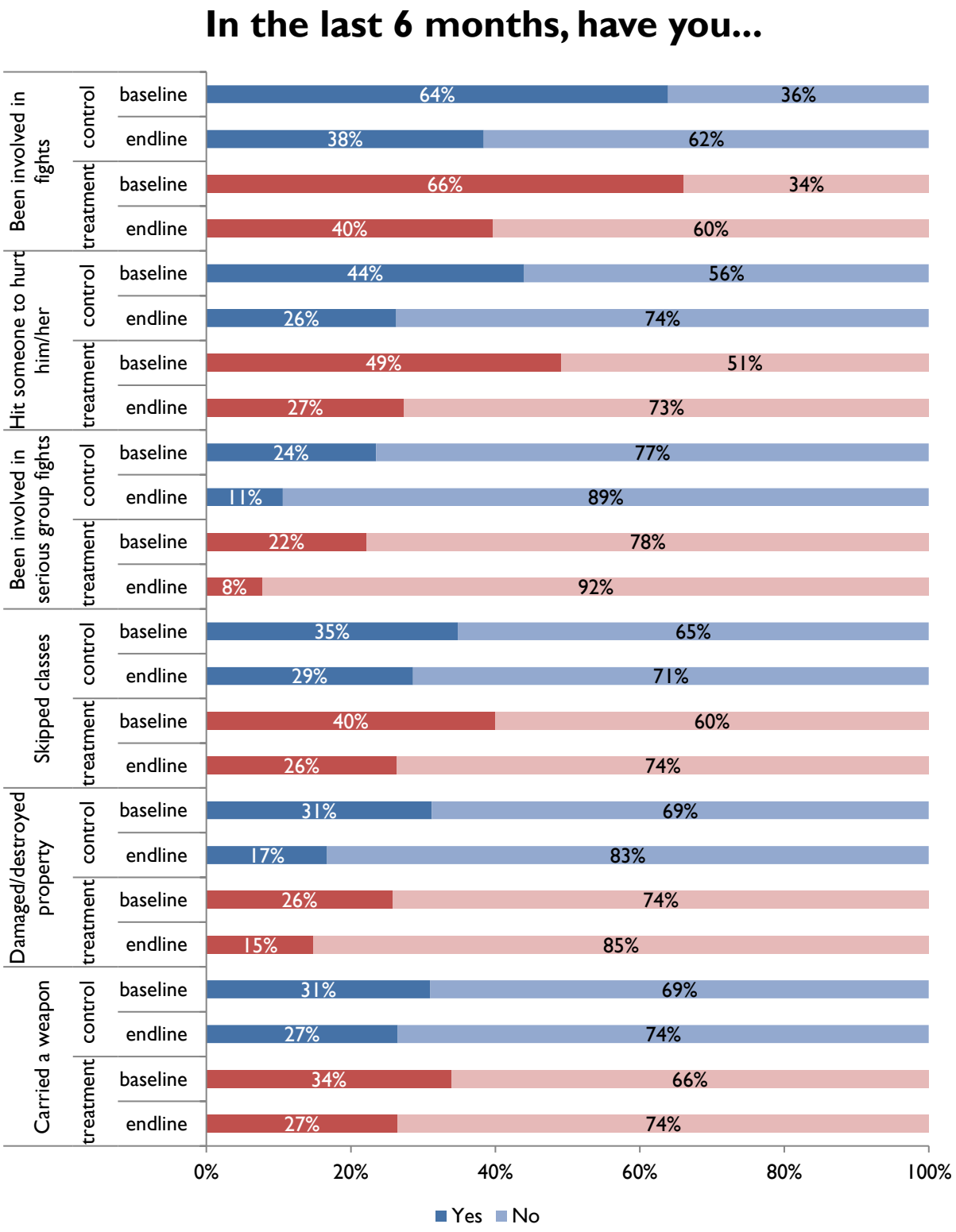


Figure 57: Share of youth responding yes to behavioral questions, less commonly reported behaviors (n=729)

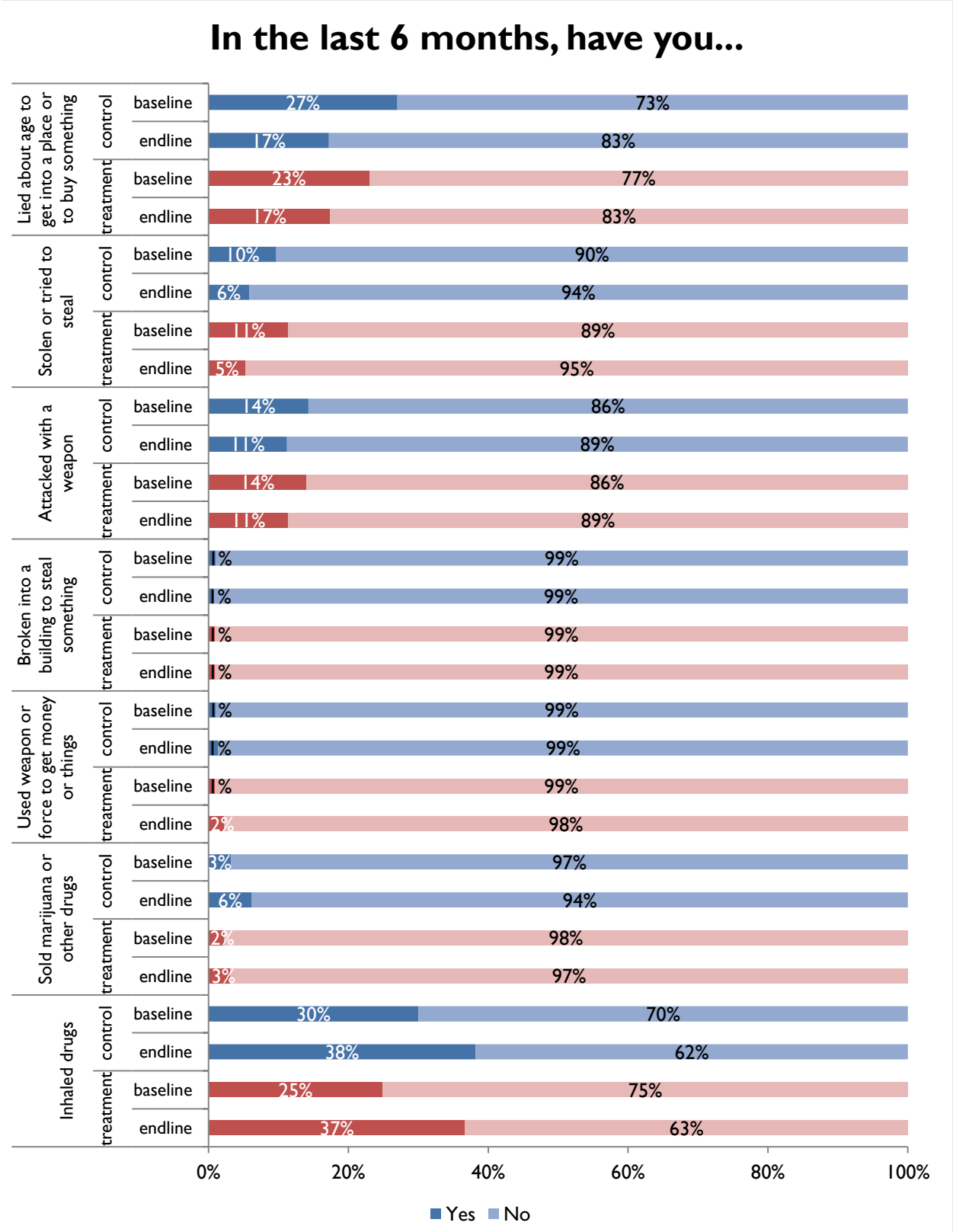


Figure 58: Questions concerning the relationship with mother/stepmother/female guardian (n=716)

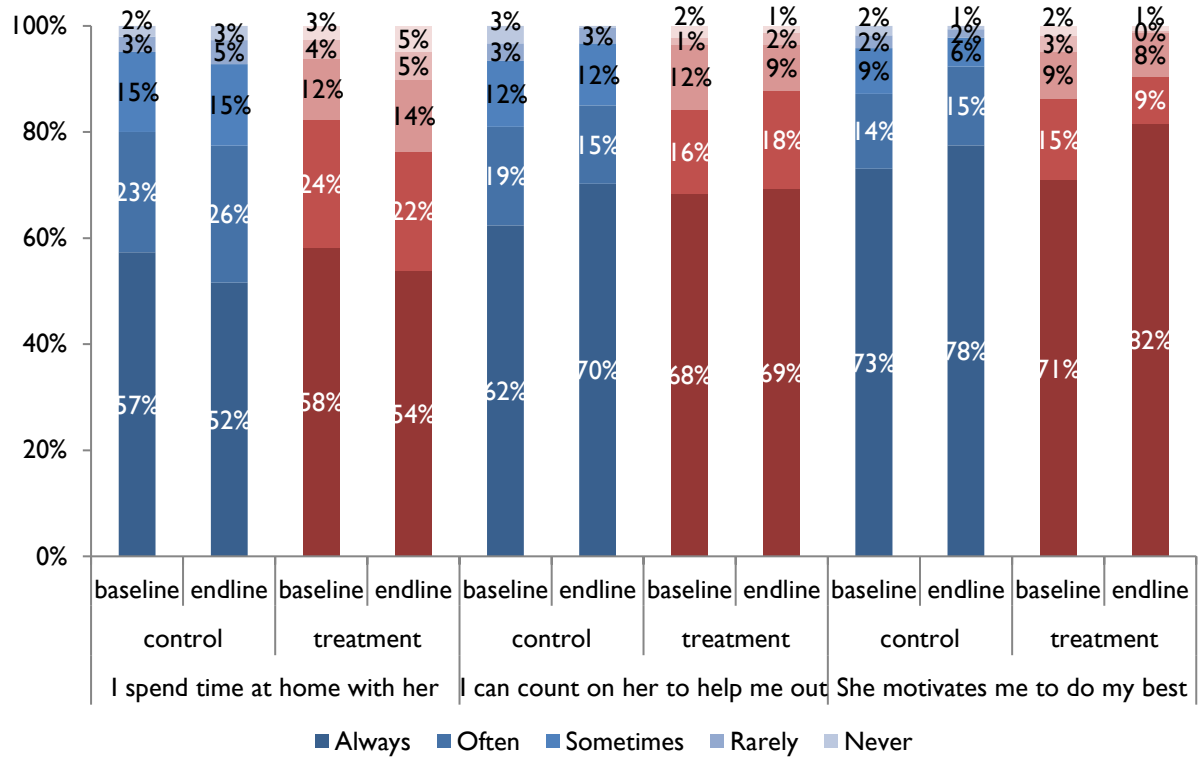


Figure 59: Questions concerning the relationship with mother/stepmother/female guardian (n=716)

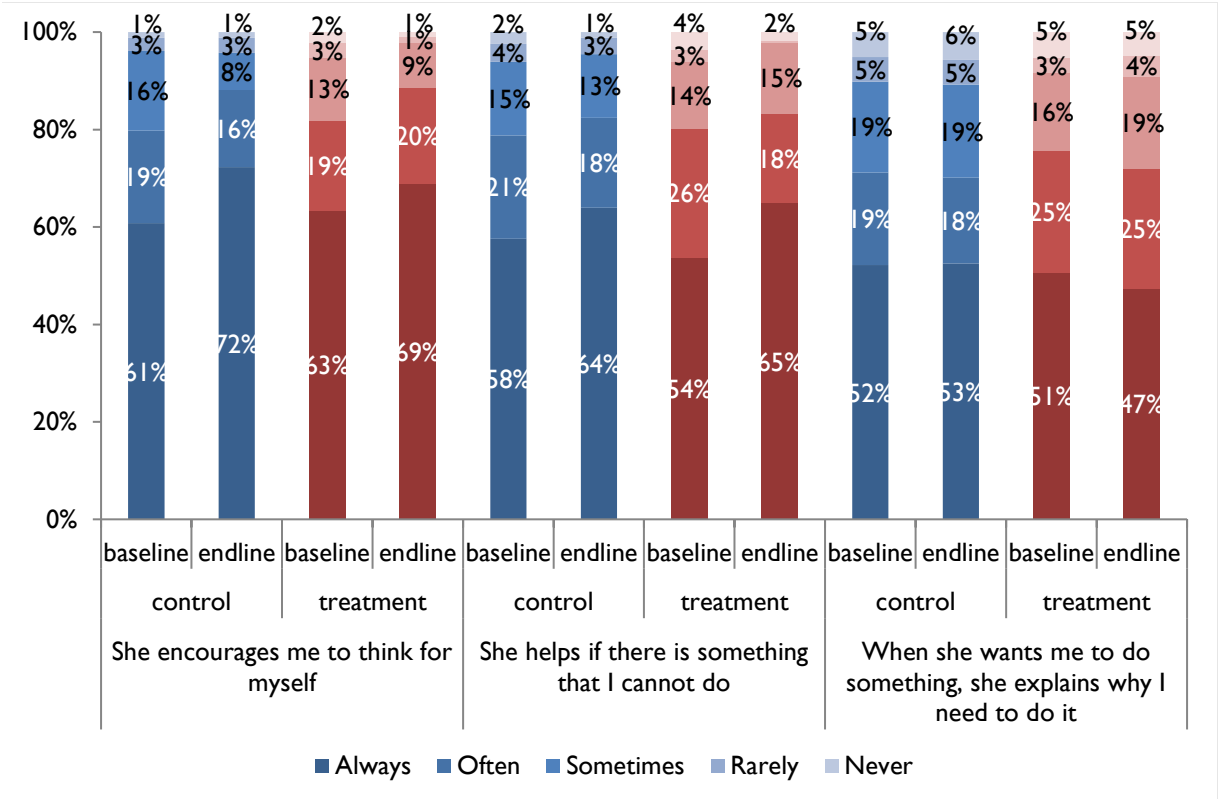


Figure 60: Module E: Frequency of communication among family members

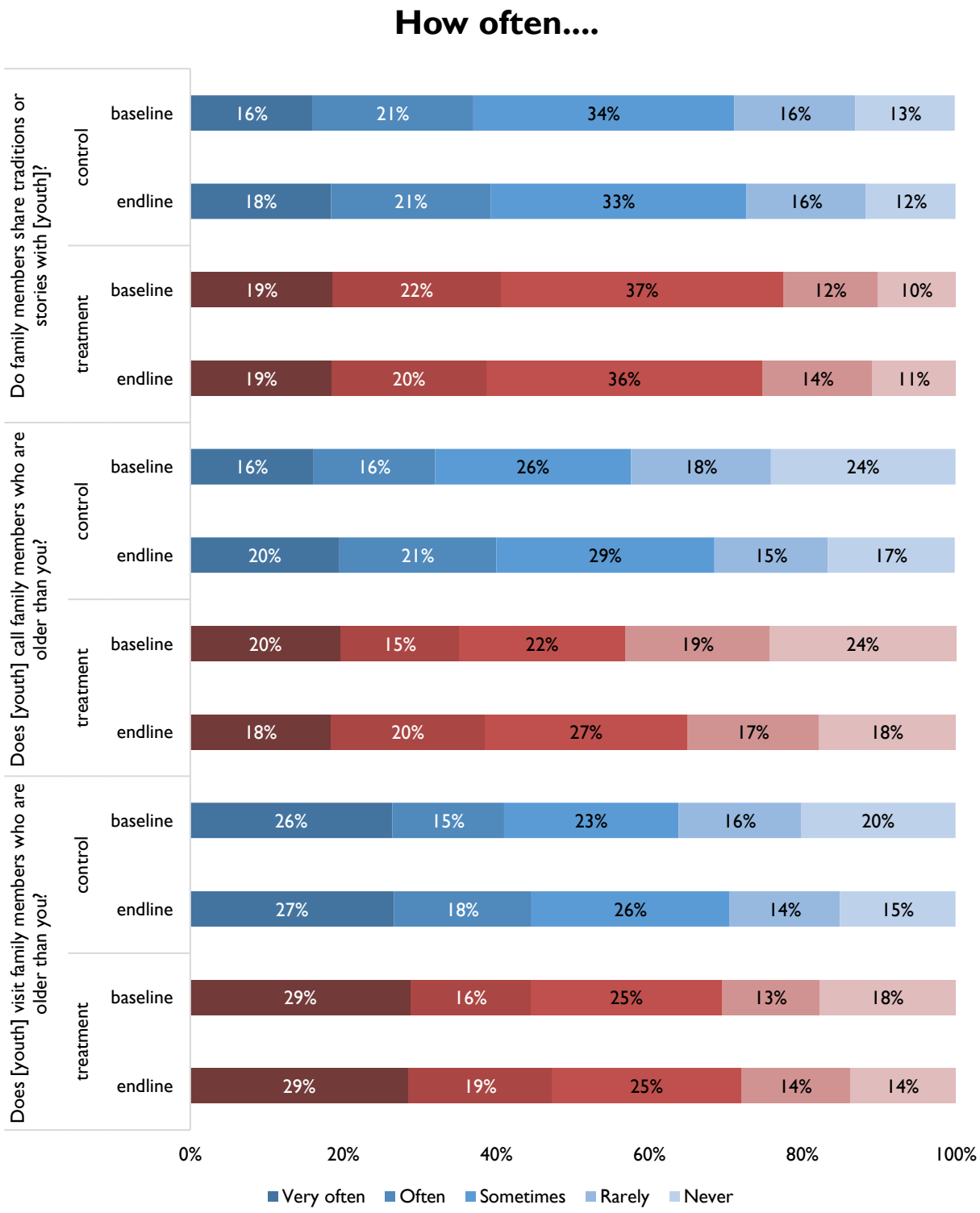


Figure 61: Module F: Caregiver behavior in front of youth

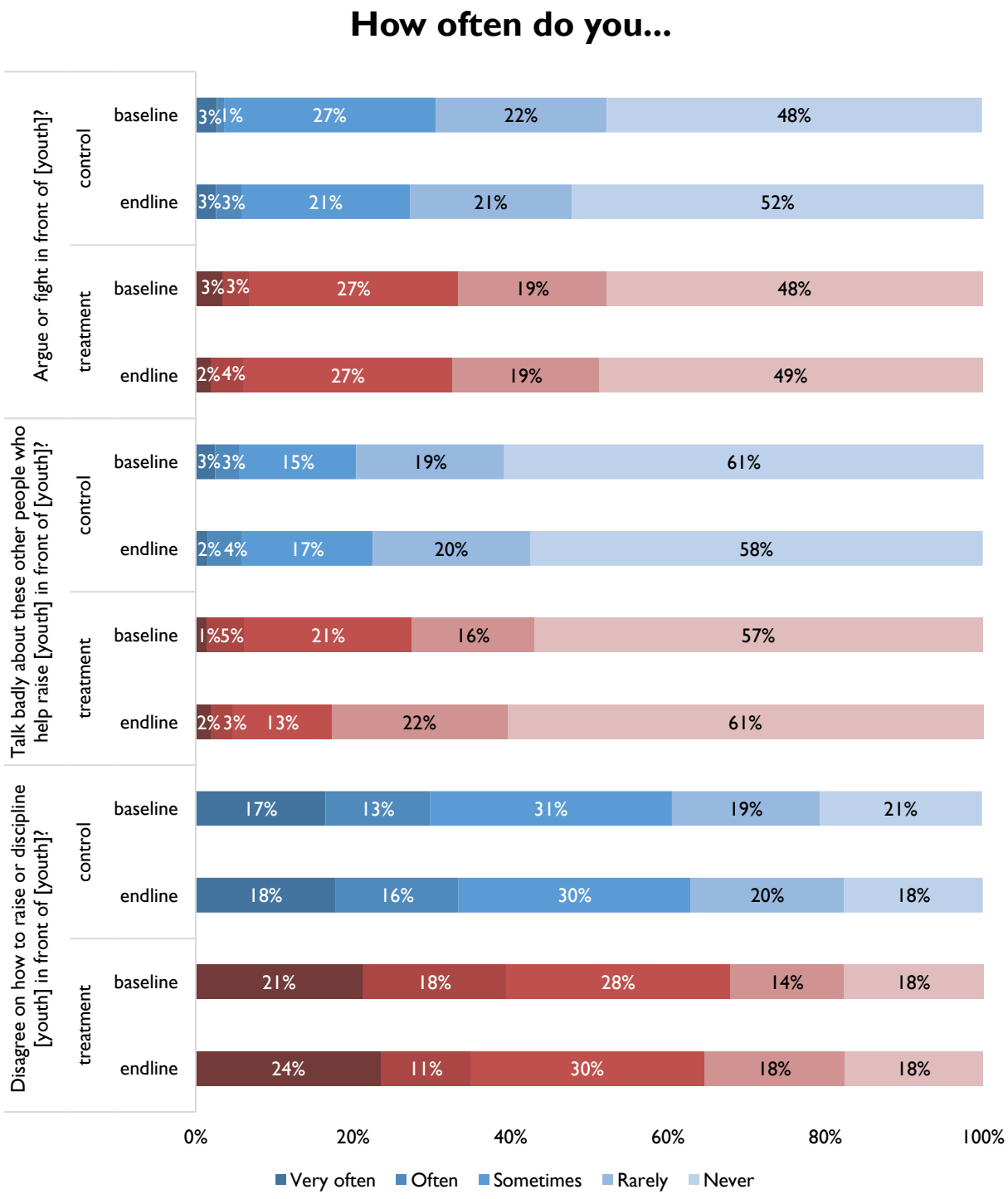


Figure 62: Module G: Caregiver agreement with different parenting philosophies

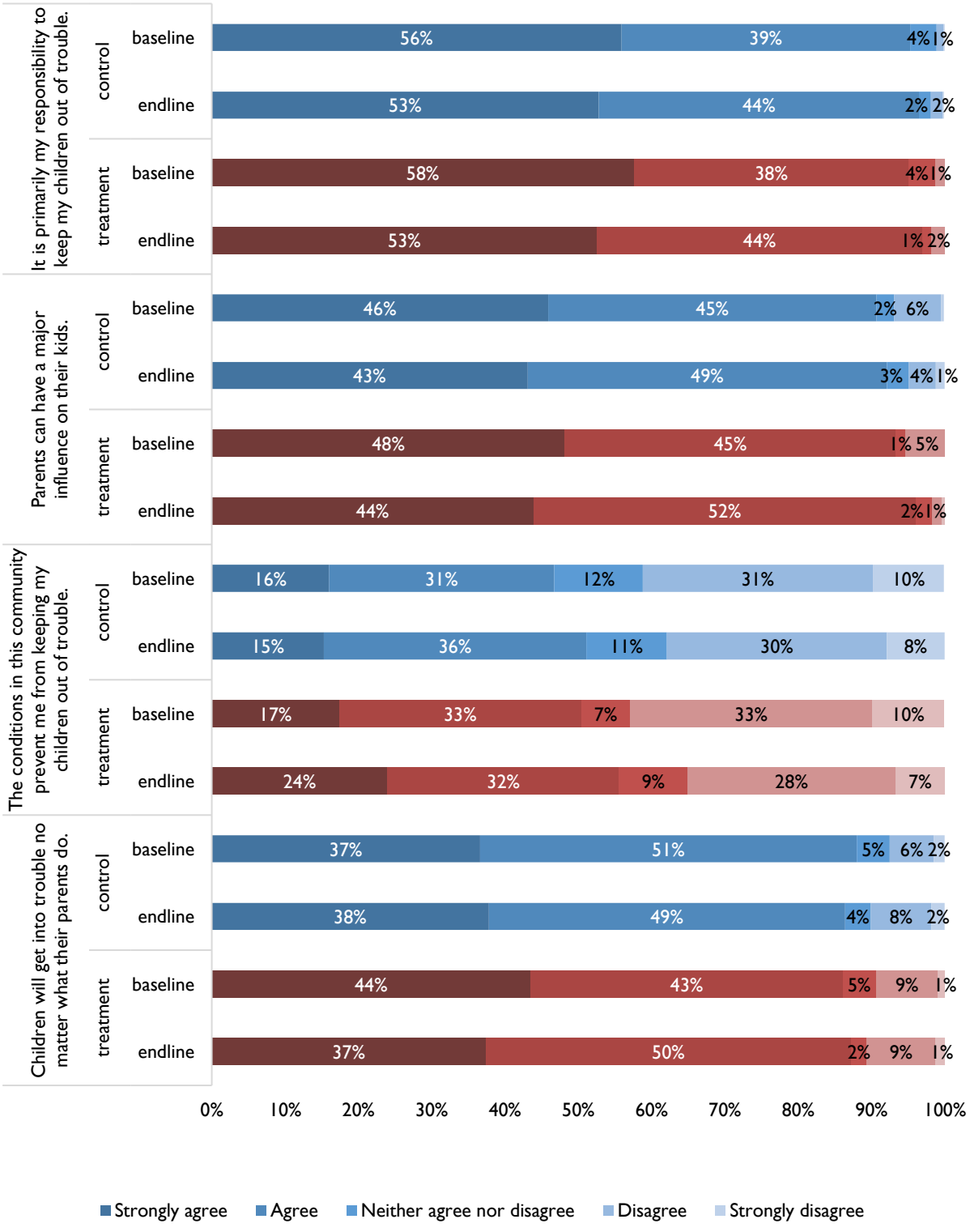


Figure 63: Module G: Self-reported changes in parenting behavior in the last 12 months

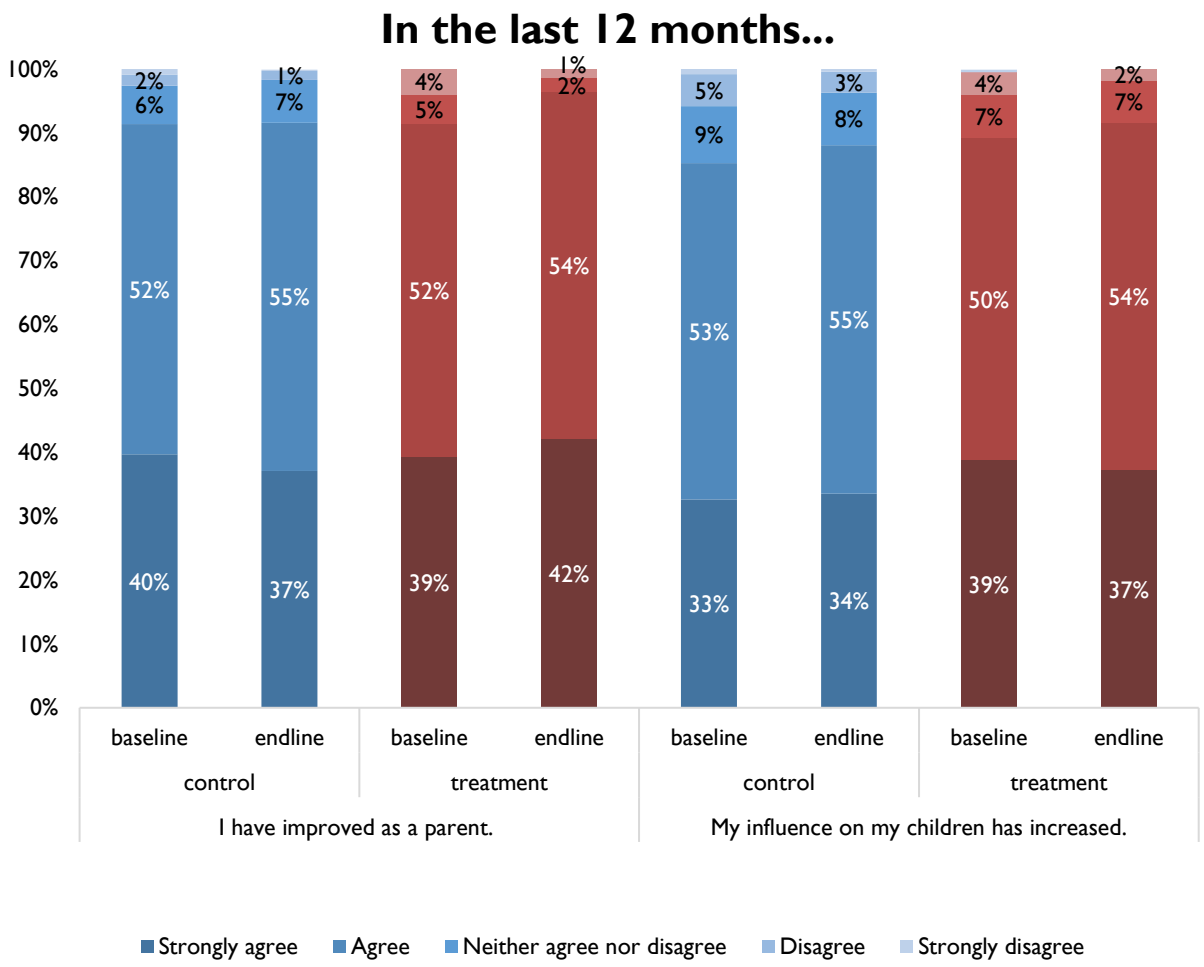


Figure 64: Module H: Reported frequency of interactions between caregiver and youth

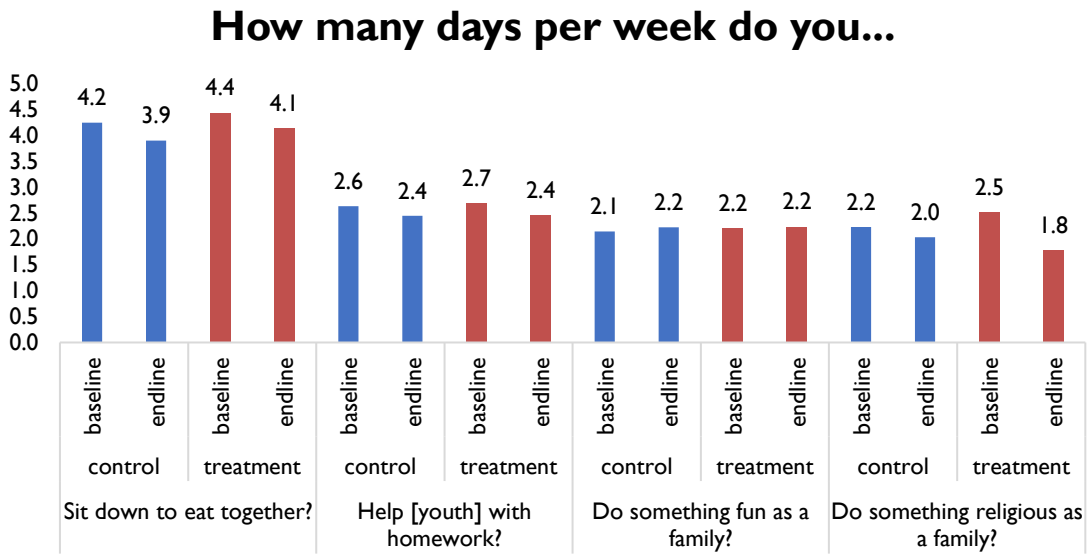


Figure 65: Module H: Caregiver's knowledge of youth whereabouts

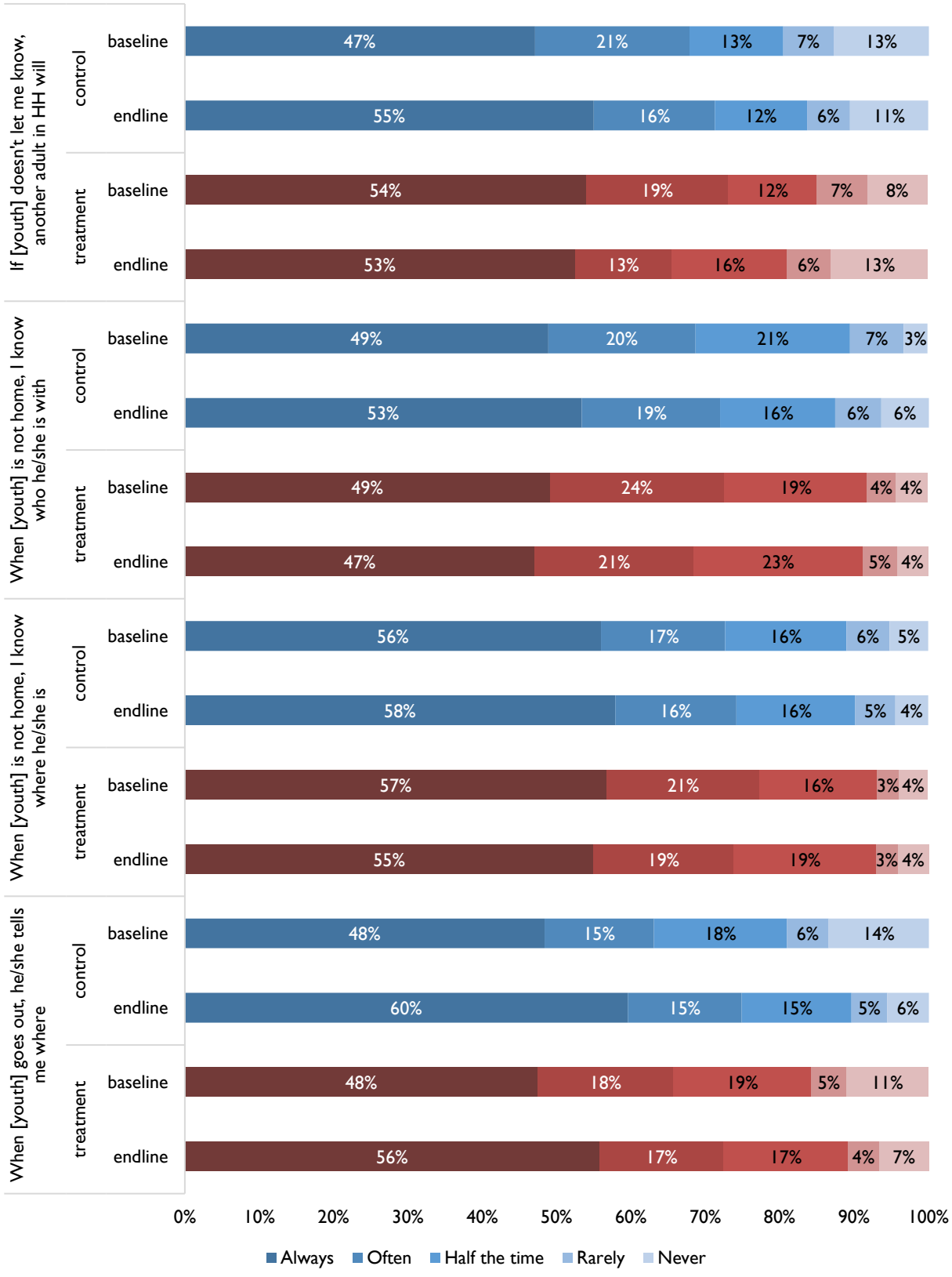


Figure 66: Module H: Communication between caregiver and youth

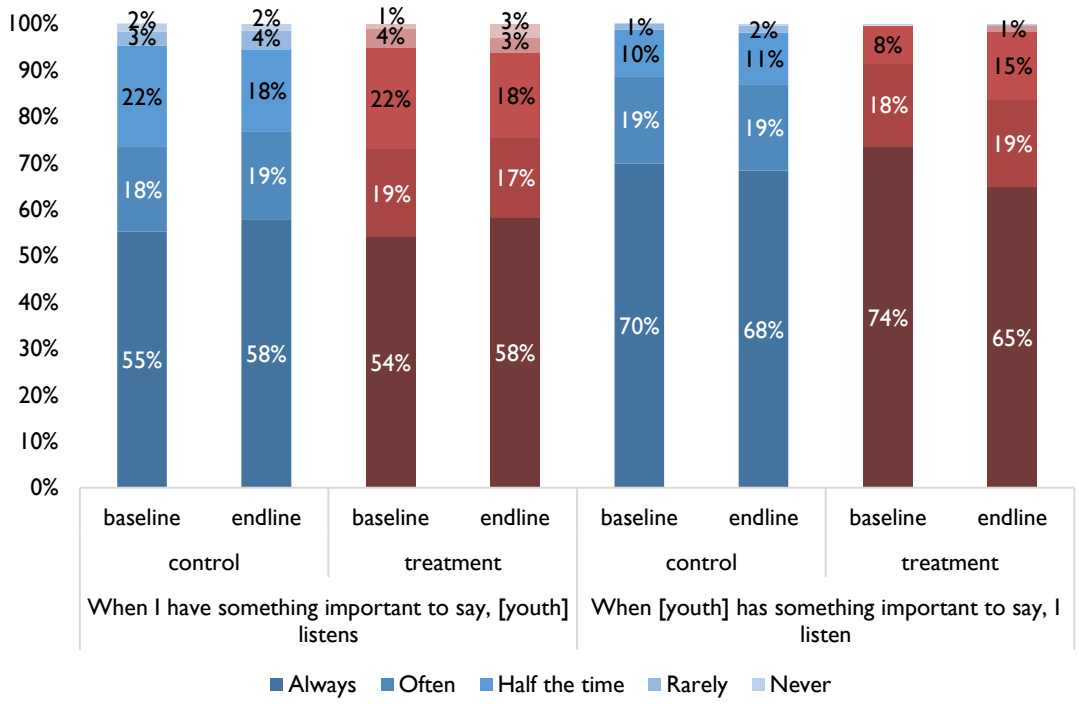


Figure 67: Module H: Reported negative interactions between caregiver and youth

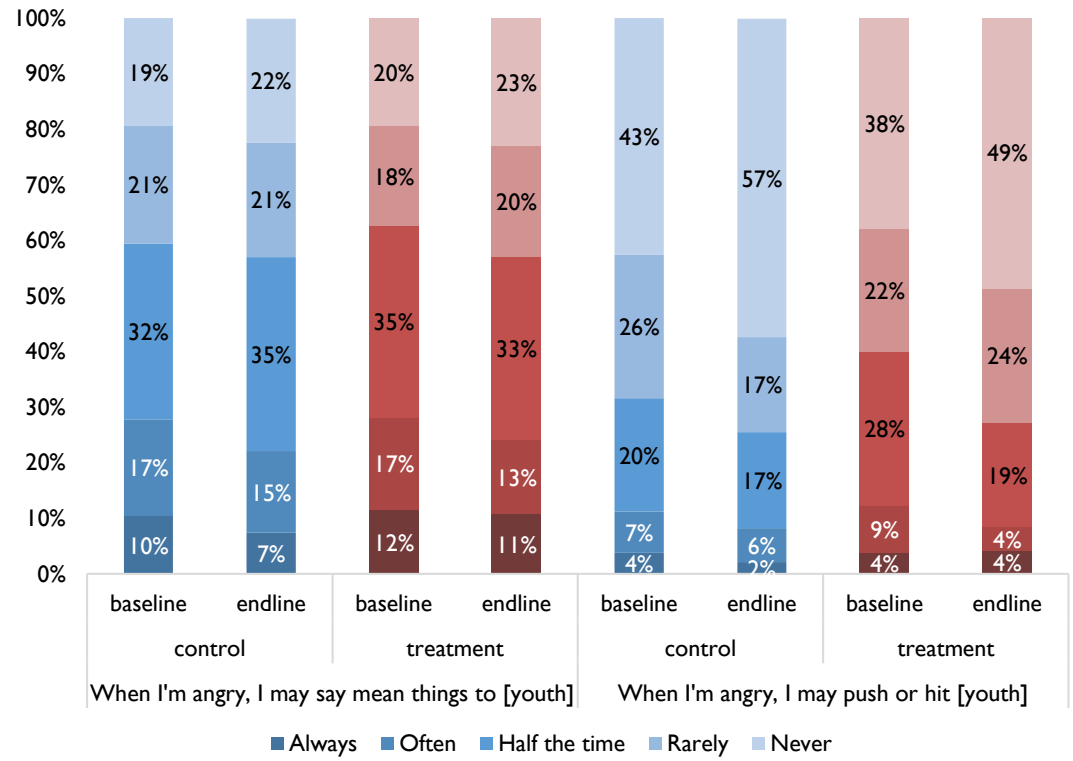


Figure 68: Module H: Reported positive interactions between caregiver and youth

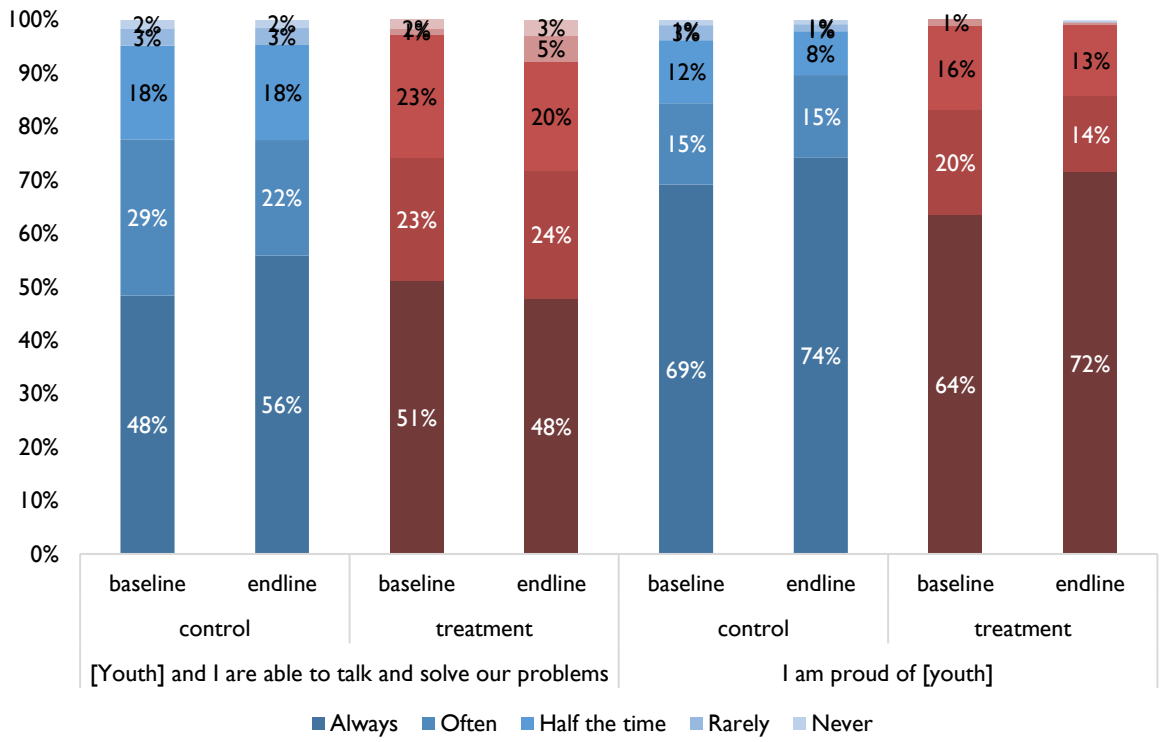


Figure 69: Module H: Household parenting choices and parenting support

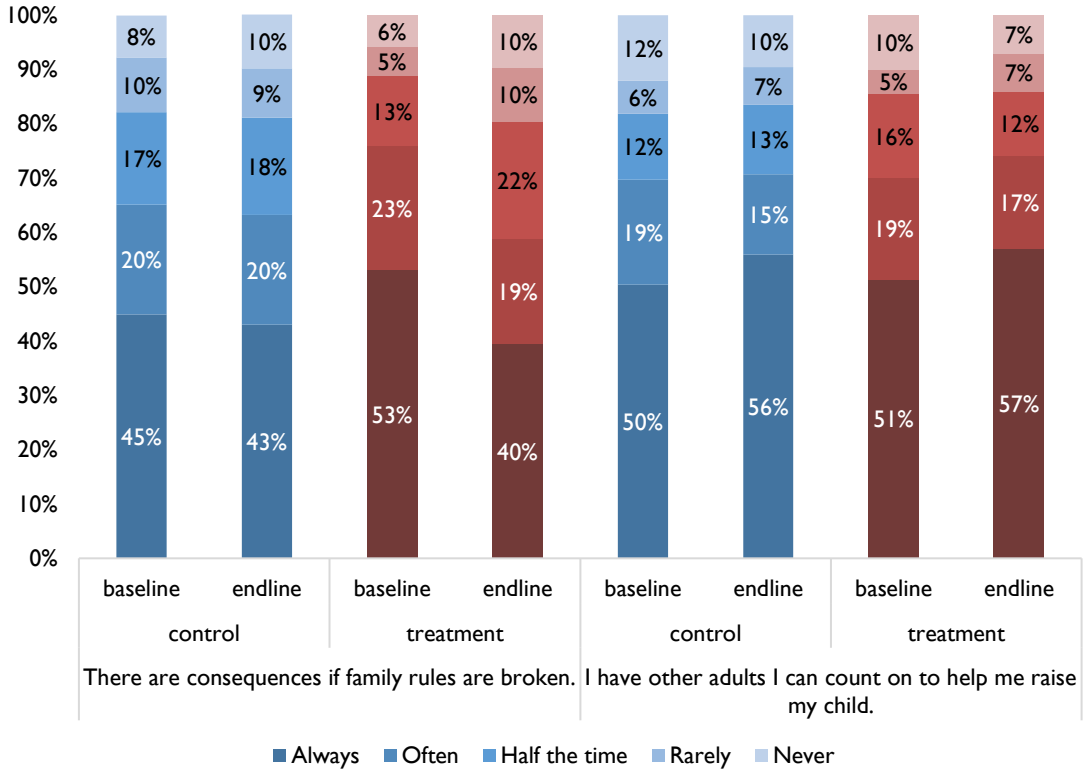


Figure 70: Module I: Youth behaviors over the last six months, as reported by caregivers

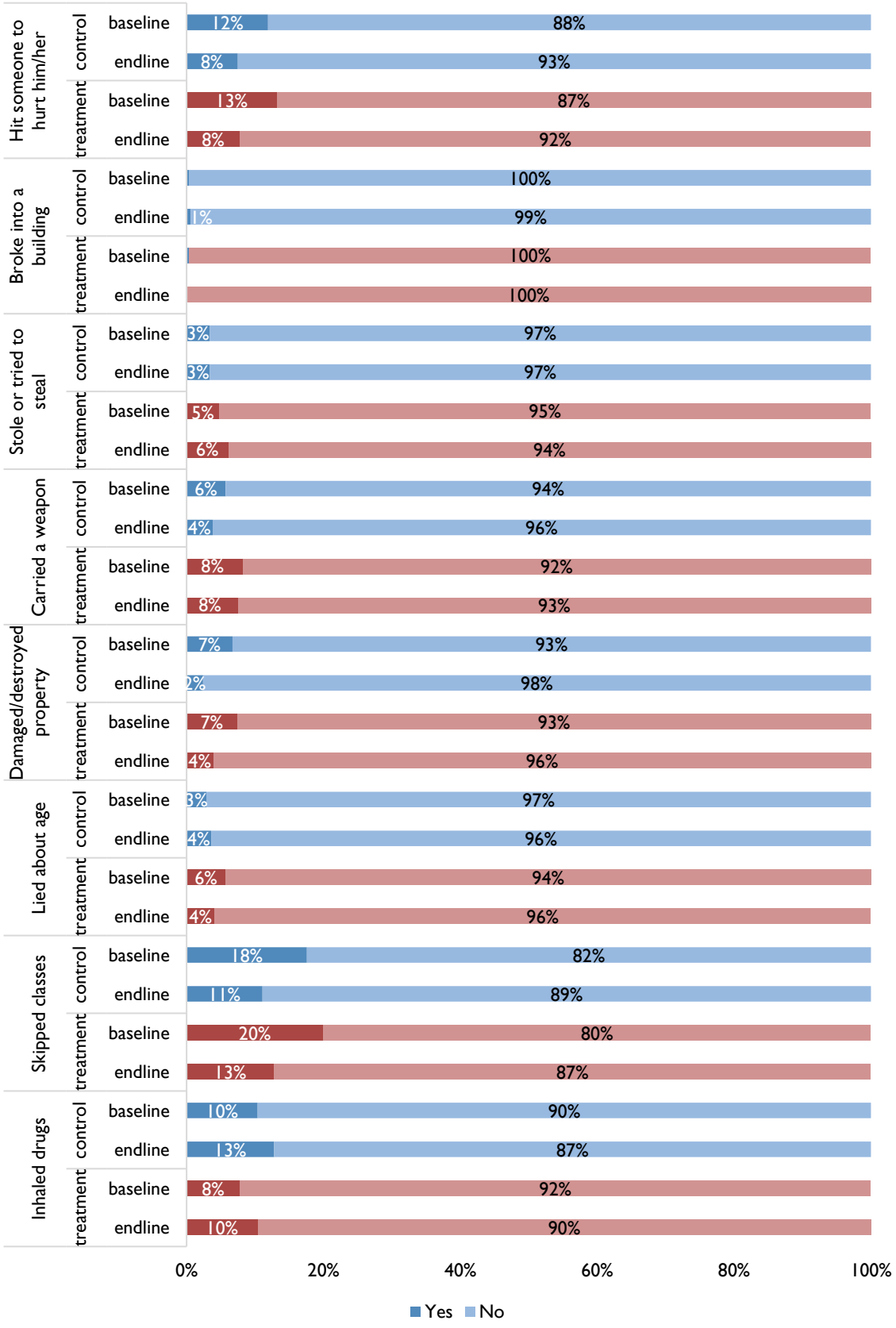


Figure 71: Module I: Youth behaviors over the last six months, as reported by caregivers (cont.)

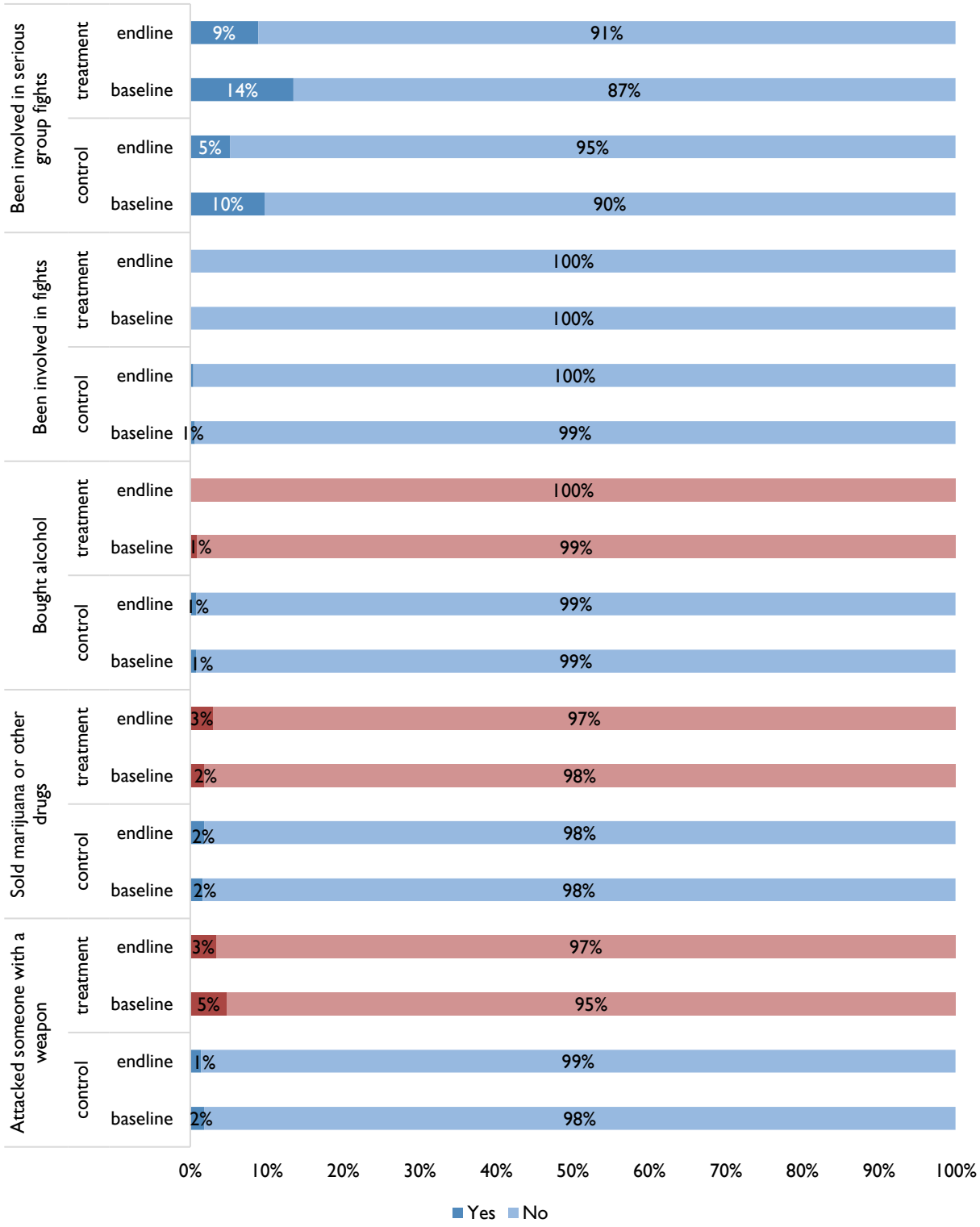


Figure 72: Module I: Caregiver discovery of youth possessing drugs

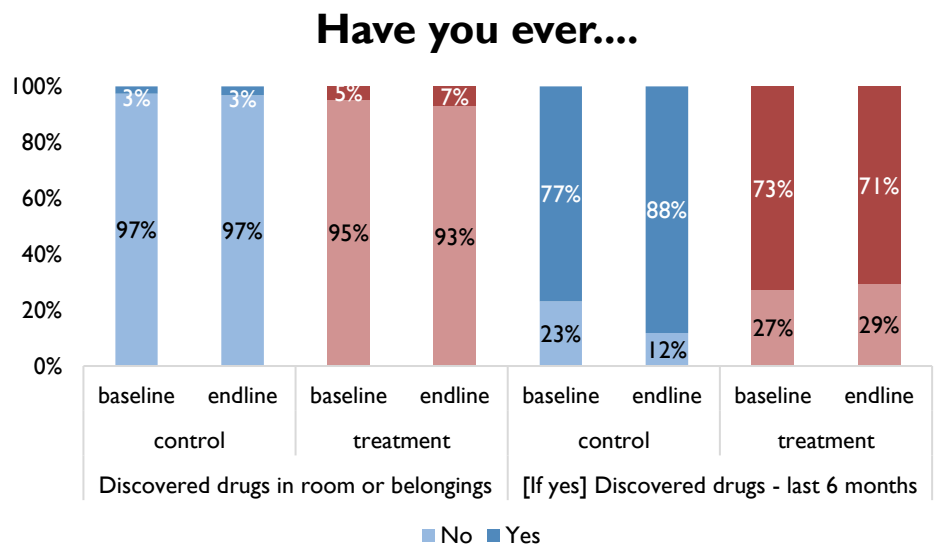


Figure 73: Module J: Reported past behaviors of other household members

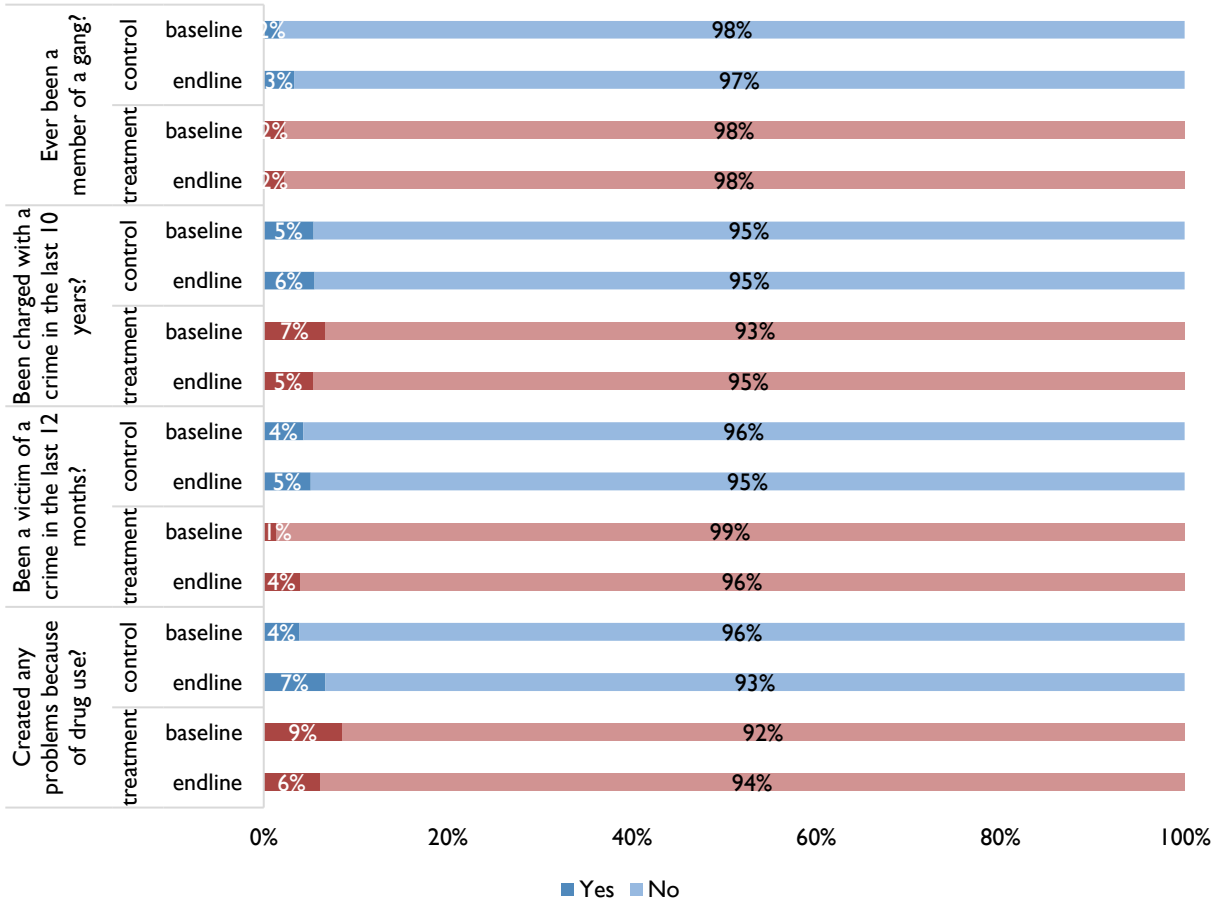


Figure 74: Module K: Amount of time caregiver reports to have lived in the community

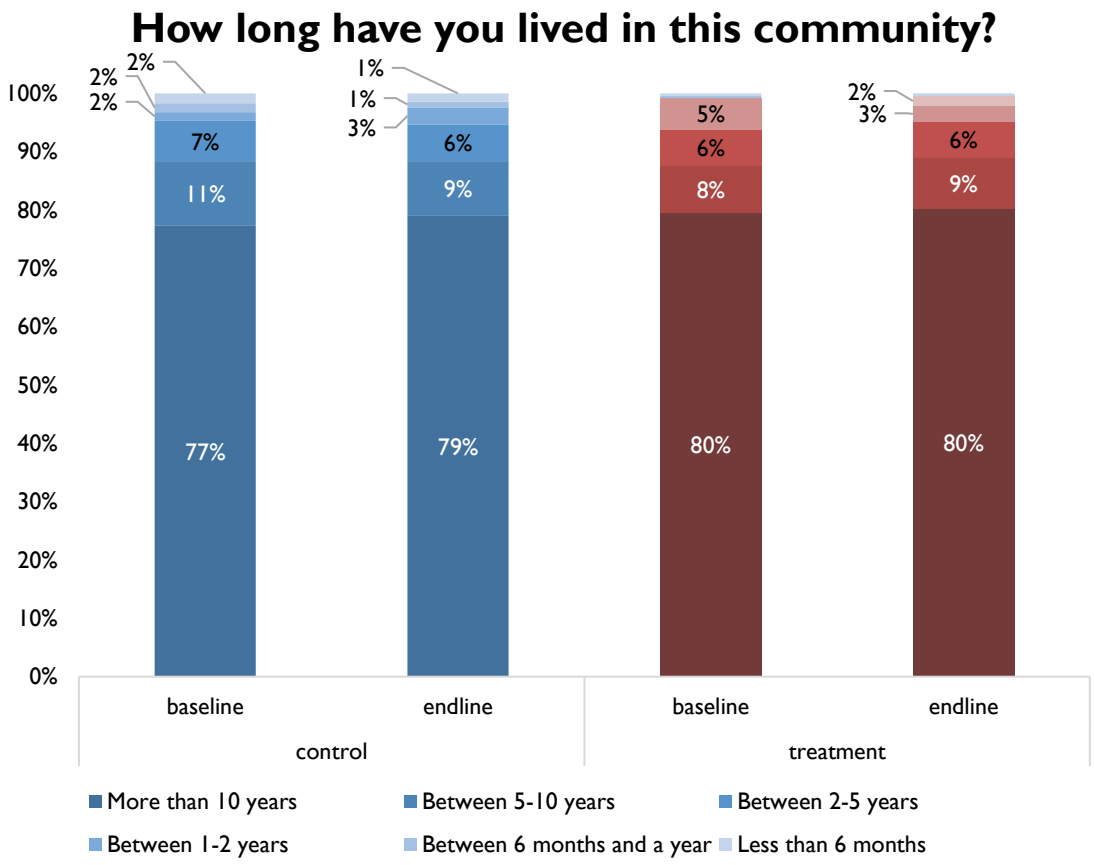


Figure 75: Module K: Relationship between caregiver and others in the neighborhood

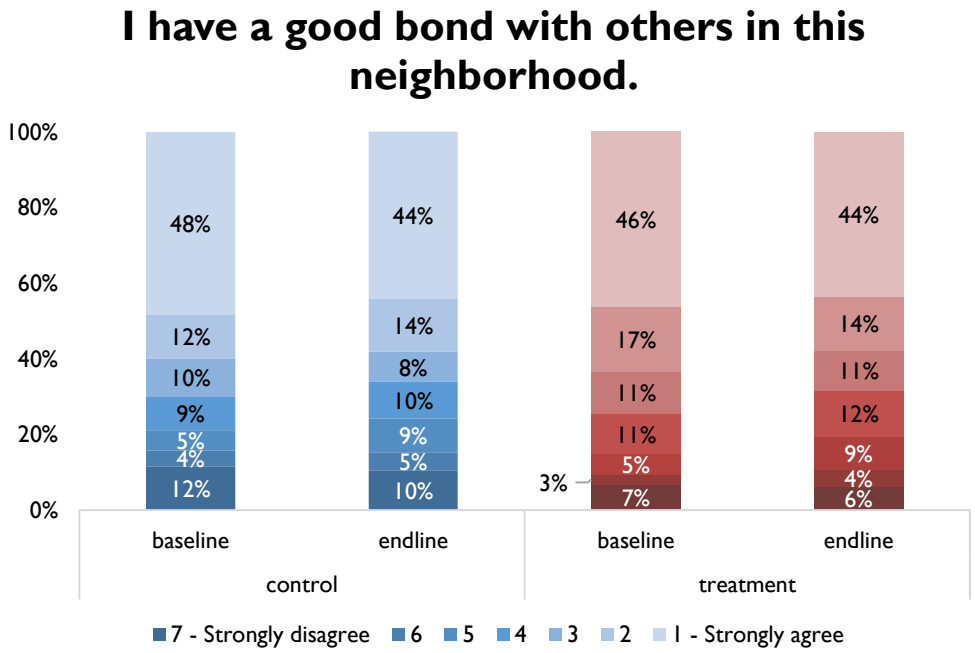


Figure 76: Module K: Caregiver satisfaction with police

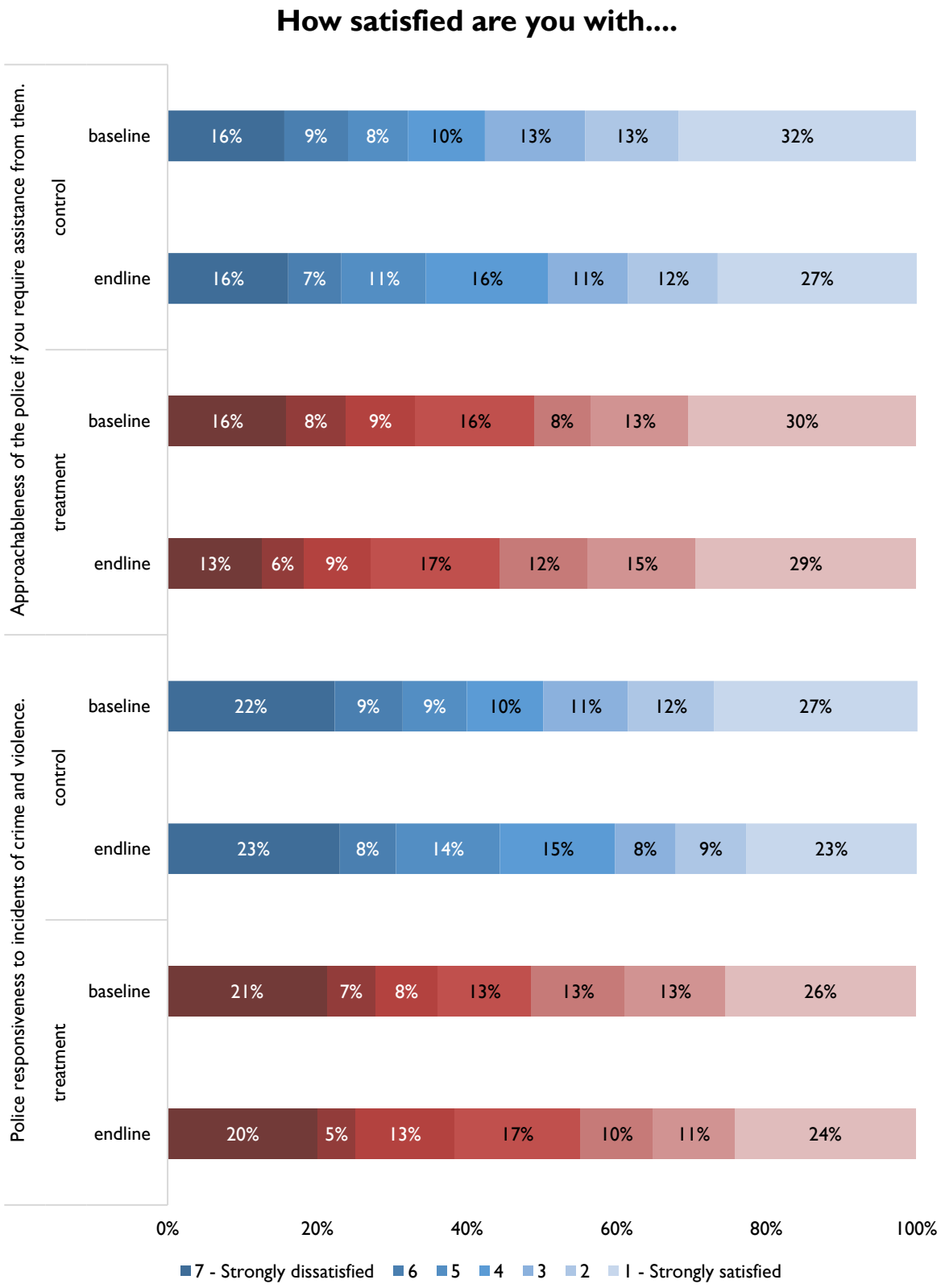


Figure 77: Module L: Reported ownership of common household items

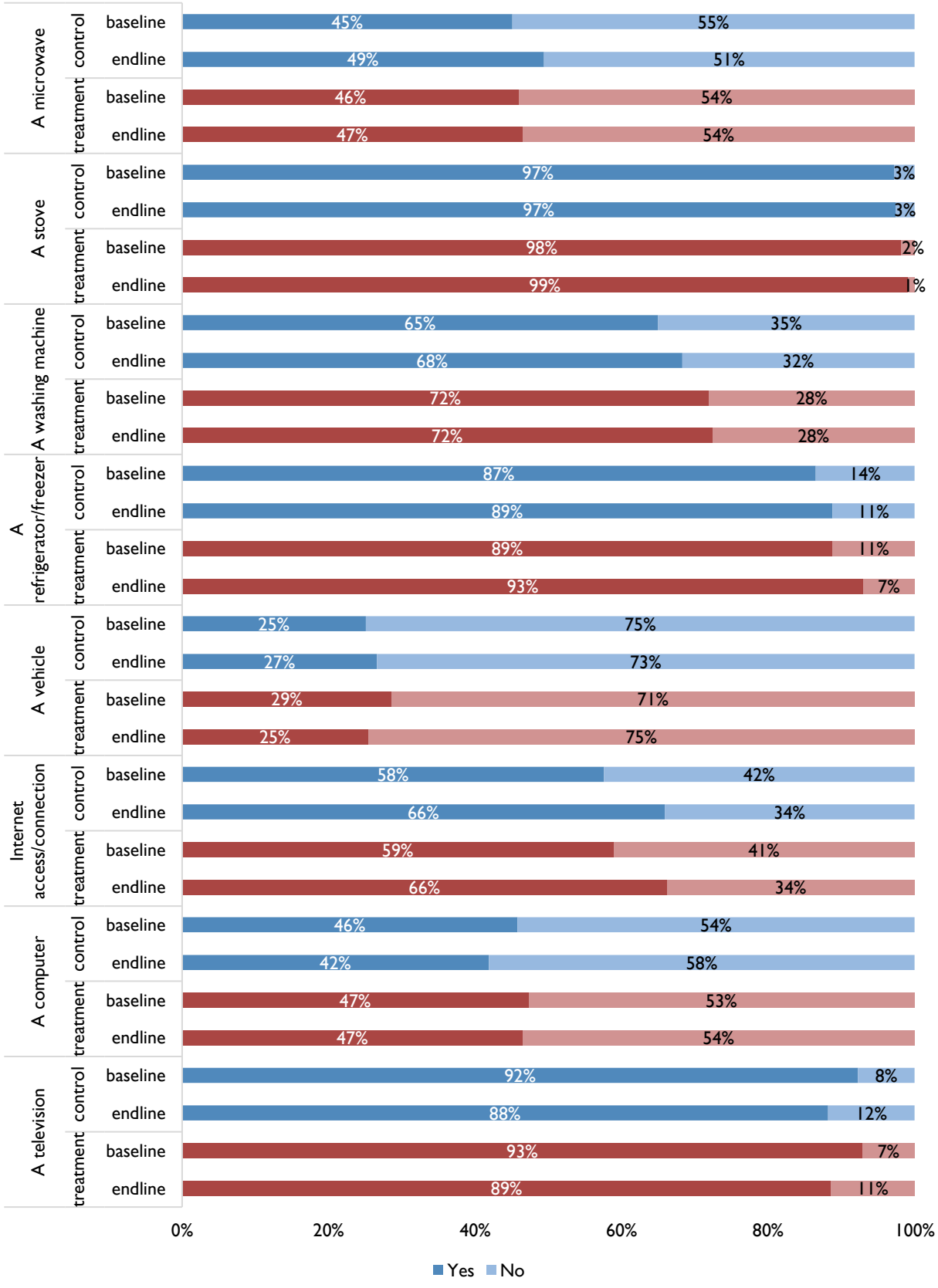


Figure 78: Module L: Reported ownership of common household items (cont.)

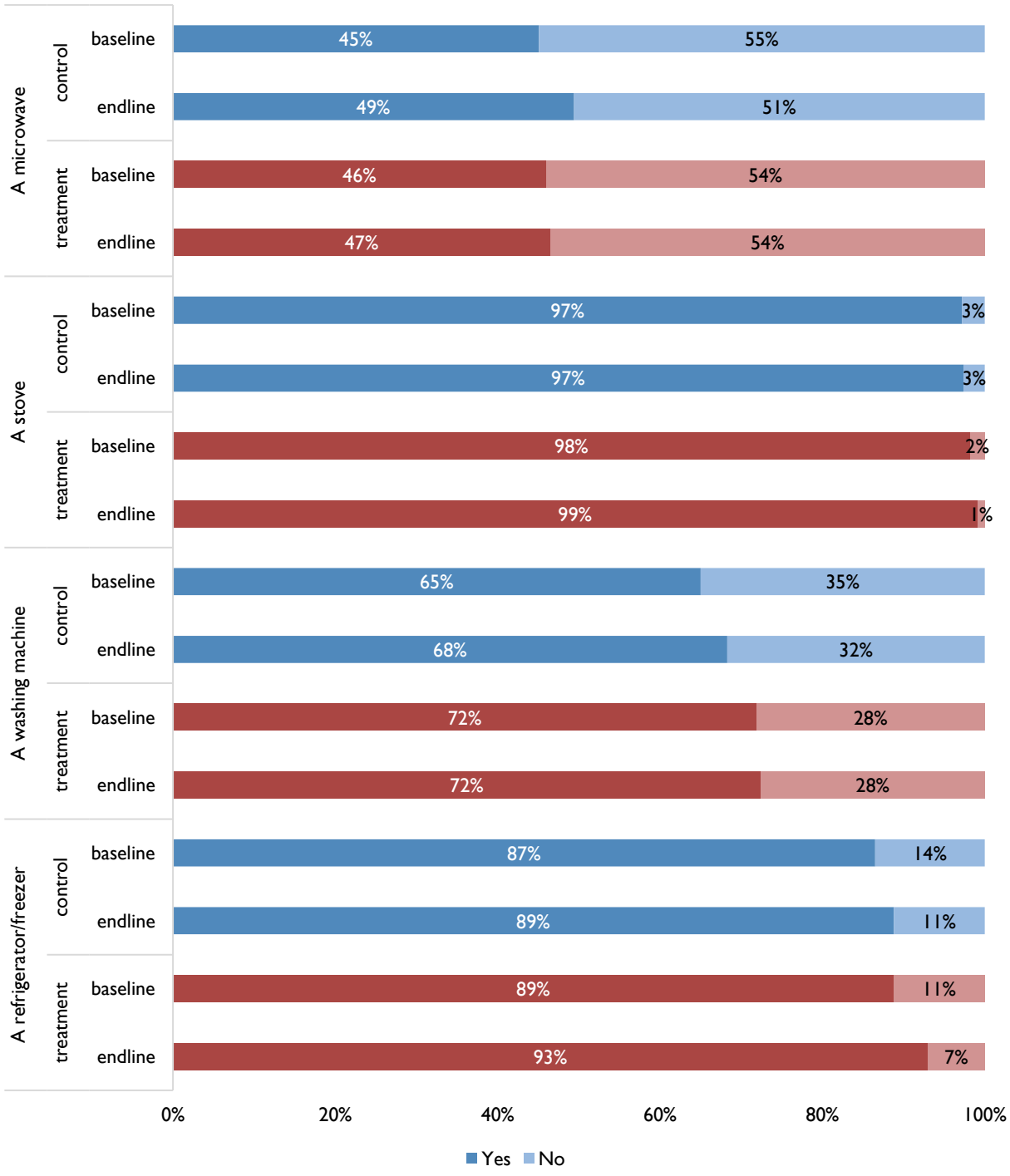
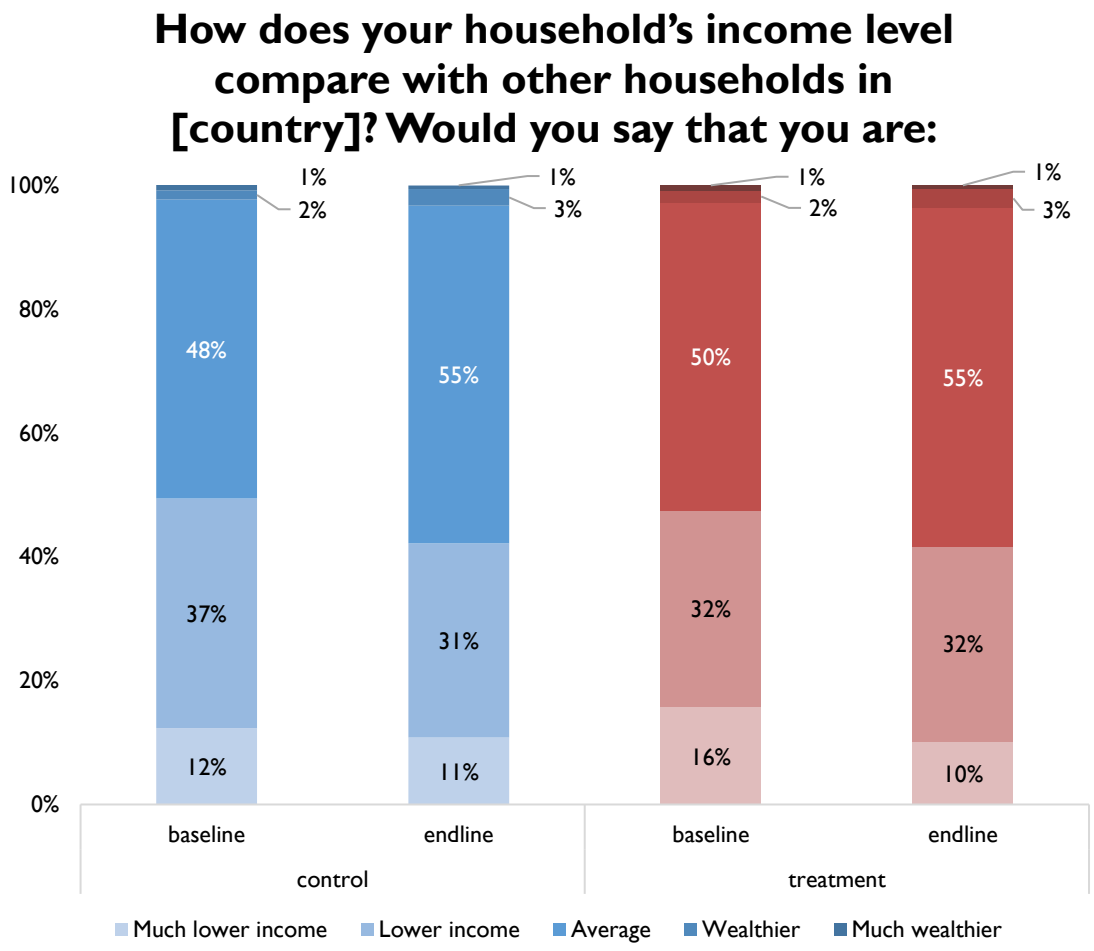


Figure 79: Module L: Reported income level relative to other households



REFERENCES

Abt, Thomas, Chris Blattman, Beatriz Magaloni, and Santiago Tobón. What Works to Prevent Violence Among Youth? A White Paper on Youth Violence, Crime Prevention, and the Mexican context. 2018.

Abt, Thomas and Christopher Winship. *What Works in Reducing Community Violence: A Meta-Review and Field Study for the Northern Triangle.* Prepared for the United States Agency for International Development (USAID). Democracy International, Inc.: Bethesda, MD, 2016. Retrieved May 2020 from https://www.usaid.gov/sites/default/files/USAID-2016-What-Works-in-Reducing-Community-Violence-Final-Report.pdf

Atienzo, Erika, Susan Baxter, and Eva Kaltenhaler. "Interventions to Prevent Youth Violence in Latin America: A Systematic Review." *International Journal of Public Health, Vol. 62, 15-29*. 2017

Atkins-Burnett, Sally. "Assessing Children and Adolescents in Large scale Surveys." January 2016. Retrieved May 2020 from https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_171794.pdf

Berk-Seligson, Susan, Diana Orcés, Georgina Pizzolitto, Mitchell Seligson, and Carole Wilson. *Impact Evaluation of USAID's Community-Based Crime and Violence Prevention* Approach in Central America: Regional Report for El Salvador, Guatemala, Honduras and Panama. Prepared for the United States Agency for International Development (USAID). The Latin American Public Opinion Project (LAPOP), Vanderbilt University: Vanderbilt, TN, 2014. Retrieved May 2020 from https://www.vanderbilt.edu/lapop/carsi/Regional_Report_v12d_final_W_120814.pdf

Betrand, Annie Melanie Beauvy-Sany, Selma Cilimkovic, Sita Conklin, and Selma Jahic. "Monitoring and Evaluation for Youth Workforce Development Projects." *The SEEP Network,* 2009. Retrieved July 2020 from https://seepnetwork.org/files/galleries/630_TN_Monitor_Eval_508.pdf

Borges, Natacha, Edith de Leeuw, and Joop Hox. "Children as Respondents in Survey Research: Cognitive Development and Response Quality." *Bulletin de Methodologie Sociologique*. April 2000. Retrieved May 2020 from http://joophox.net/publist/bms66.pdf

Brathwaite, Marjorie, Carlyle Glean, Vere Goodridge, Stephenson Grayson, and Sybil Leslie. *The Caribbean: Our Land and People: Book 1*. Oxford: Heinemann International, 1988.

Cahill, Meagan, Jesse Jannetta, Emily Tiry, Samantha Lowry, Miriam Becker-Cohen, Ellen Paddock, Maria Serakos, Loraine Park, and Karen Hennigan. *Evaluation of the Los Angeles Gang Reduction and Youth Development Program: Year 4 Evaluation Report.* Urban Institute and Harder+ Company Community Research, 2015. Retrieved May 2020 from https://www.urban.org/sites/default/files/publication/77956/2000622-Evaluation-of-the-Los-Angeles-Gang-Reduction-and-Youth-Development-Program-Year-4-Evaluation-Report.pdf

Camerini, A., Schulz, P.J. "Social Desirability Bias in Child-Report Social Well-Being: Evaluation of the Children's Social Desirability Short Scale Using Item Response Theory and Examination of Its Impact on Self-Report Family and Peer Relationships." *Child Ind Res* **11,** 2018. Retrieved July 2020 from https://link.springer.com/article/10.1007/s12187-017-9472-9

Campie, Patricia, Manolya Tanyu, and Chinmaya Udayakuma Holla *What Works to Prevent Lethal Youth Violence in the LAC Region: A Global Review of the Research.* Washington, DC: American Institutes for Research, 2019. Retrieved May 2020 from https://www.youthpower.org/sites/default/files/YouthPower/files/resources/Evidence-Mapping-Report-508.pdf

CARICOM Commission on Youth Development (CCYD). *Eye on Future: Investing in YOUTH NOW for Tomorrow's Community.* 2010. Retrieved May 2020 from https://www.youthjamaica.com/sites/default/files/CCYD%20Report.pdf

Caribbean Development Bank. "2019 Caribbean Economic Review and 2020 Outlook." Retrieved May 2020 from https://www.caribank.org/publications-and-resources/resource-library/infographics/2019-caribbean-economic-review-and-2020-outlook

Chioda, Laura. *Stop the Violence in Latin America: A Look at Prevention from Cradle to Adulthood.* Washington, DC: World Bank Group, 2017. Retrieved May 2020 from https://doi.org/10.1596/978-1-4648-0664-3

Chu, Chi Meng, Michael Daffern, Stuart Thomas, and Jia Ying Lim. "Elucidating the treatment needs of gang-affiliated youth offenders." *Journal of Aggression, Conflict and Peace Research, Vol. 3 Iss:3, 129-140.* 2009. Retrieved May 2020 from https://doi.org/10.1108/17596591111154167

CFYR. 2017. Family Counselor Training Manual – Draft. Family Counselor Training - St Lucia. November 27 – December 1, 2017.

Creative Associates International. "Building Resilient Youth, Tunisia // ETTYSAL." *Dispatches: Think Creative, Fall 2018, 10-11.* 2018.

Creative Associates International. *Community, Family, and Youth Resilience (CFYR) Program: Performance Monitoring Plan (PMP).* Prepared for the United States Agency for International Development (USAID). 2017.

Creative Associates International. *Sistematización de la experiencia piloto del Programa de Prevención y Oportunidades: Trabajando para el Entendimiento Familiar (PROPONTE).* Prepared for the United States Agency for International Development (USAID). 2015.

De Leeuw, Edith. "Improving Data Quality when Surveying Children and Adolescents: Cognitive and Social Development and its Role in Questionnaire Construction and Pretesting." May 2011. Retrieved May 2020 from https://www.aka.fi/globalassets/awanhat/documents/tiedostot/lapset/presentations-of-the-annual-seminar-10-12-may-2011/surveying-children-and-adolescents_de-leeuw.pdf

Dunworth, Terence, Dave Hayeslip, Morgan Lyons, and Megan Denver. *Evaluation of the Los Angeles Gang Reduction and Youth Development Program: Final Y1 Report.* The Urban Institute and Harder+ Company Community Research, 2010. Retrieved May 2020 from https://www.urban.org/sites/default/files/publication/29326/412251-Evaluation-of-the-Los-Angeles-Gang-Reduction-and-Youth-Development-Program-Final-Y-Report.PDF

Dunworth, Terence, David Hayeslip, and Megan Denver. *Y2 Final Report: Evaluation of the Los Angeles Gang Reduction and Youth Development Program.* The Urban Institute and Harder+ Company

Community Research, 2011. Retrieved May 2020 from https://www.urban.org/sites/default/files/publication/27581/412409-Y-Final-Report-Evaluation-of-the-Los-Angeles-Gang-Reduction-and-Youth-Development.PDF

Exxon Mobil. "Guyana project overview." February 5, 2020. Retrieved May 2020 from https://corporate.exxonmobil.com/Locations/Guyana/Guyana-project-overview#DiscoveriesintheStabroekBlock

Foss, Aaron, Bertrand Laurent, Danielle de Garcia, Henry Charles, Mateusz Pucilowski, Nicole Hazel, Rajwantie Sahai, and Sharene McKenzie. Eastern and Southern Caribbean Youth Assessment (ESCYA): Final Report. Prepared for the United States Agency for International Development (USAID). Social Impact, Inc: Arlington, VA, 2013. Retrieved May 2020 from http://www.undp.org/content/dam/trinidad_tobago/docs/DemocraticGovernance/Publications/ESCYA%20Final%20Report%20(FINAL%20SUBMISSION)%2031%20Oct.pdf

Fuchs, Marek. "Children and Adolescents as Respondents. Experiments on Question Order, Response Order, Scale Effects and the Effect of Numeric Values Associated with Response Options." 2005. Retrieved May 2020 from https://pdfs.semanticscholar.org/d576/862830357a1c9282d161bda239bfb7db679f.pdf

Fuchs, Marek. "The Reliability of Children's Survey Responses: The Impact of Cognitive Functioning on Respondent Behavior." 2009. Retrieved May 2020 from https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2008000/article/10961-eng.pdf?st=CfumMDNm

Galesic, Mirta and Bosnjak, Michael. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly*, 2009. Retrieved July 2020 from https://www.researchgate.net/publication/30967227_Effects_of_Questionnaire_Length_on_Participation_and_Indicators_of_Response_Quality_in_a_Web_Survey

Ghanem, Dalia Sarojini Hirshleifer and Karen Ortiz-Becerra. "Testing Attrition Bias in Field Experiments." *Working Papers 201919.* University of California at Riverside, Department of Economics, 2019. Retrieved May 2020 from https://economics.ucr.edu/repec/ucr/wpaper/201919.pdf

Guerra, Nancy, Kirk Williams, Ian Walker and Julie Meeks-Gardner. "Building an Ecology of Peace in Jamaica: New Approaches to Understanding Youth Crime and Violence and Evaluating Prevention Strategies." World Bank Technical Report. 2013.

Haushofer, Johannes and Jeremy Shapiro. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *The Quarterly Journal of Economics, 1973-2042.* 2016.

Henggeler, Scott and Ashli Sheidow. "Empirically Supported Family-Based Treatment for Conduct Disorder and Delinquency in Adolescents." *Journal of Marital and Family Therapy.* 2011. Retrieved May 2020 from https://doi.org/10.1111/j.1752-0606.2011.00244.x

Hellevik, Ottar. "Linear versus logistic regression when the dependent variable is a dichotomy." *Qual Quant* 43, 59-74. 2009. Retrieved May 2020 from https://link.springer.com/article/10.1007/s11135-007-9077-3

Hennigan, Karen, Cheryl Maxson, David Sloane, Kathy Kolnick, and Flor Vindel. "Identifying high-risk youth for secondary gang prevention." *Journal of Crime and Justice, 37:1, 104-128.* 2014. Retrieved May 2020 from https://doi.org/10.1080/0735648X.2013.831208

Hennigan, Karen, Kathy Kolnick, Flor Vindel, and Cheryl Maxson. "Targeting youth at risk for gang involvement: Validation of a gang risk assessment to support individualized secondary preventions." *Children and Youth Services Review, Vol. 56, 86-96*. 2015. Retrieved May 2020 from https://doi.org/10.1016/j.childyouth.2015.07.002

Herzog, A. Regula and Bachman, Jerald G. "Effects of Questionnaire Length on Response Quality." *Public Opinion Quarterly, Vol. 45, Issue 4, Winter 1981*. Retrieved July 2020 from https://academic.oup.com/poq/article-abstract/45/4/549/1849451?redirectedFrom=fulltext

Jaitman, Laura and Roberto Guerrero Compeán. *Closing Knowledge Gaps: Toward Evidence-Based Crime Prevention Policies in Latin America and the Caribbean.* Washington, DC: Inter-American Development Bank, 2015. Retrieved May 2020 from https://publications.iadb.org/publications/english/document/Closing-Knowledge-Gaps-Toward-Evidence-Based-Crime-Prevention-Policies-in-Latin-America-and-the-Caribbean.pdf

Katz, Charles, Hyunjung Cheon, and Scott Decker. *An Evaluation of Proponte Más: a Honduran Secondary Prevention Program Technical Report.* Prepared for the United States Agency for International Development (USAID). Arizona State University: Phoenix, Arizona, 2019

Katz, Charles and Lidia Nuño. *Prevalence and Patterns of Troublesome Youth Groups in the Caribbean: Final Report.* Prepared for the United States Agency for International Development (USAID). Arizona State University: Phoenix, Arizona, 2017.

Katz, Charles and Lidia Nuño. *Prevalence and Patterns of Troublesome Youth Groups in the Caribbean: Implications for Policy and Practice.* Prepared for the United States Agency for International Development (USAID). Arizona State University: Phoenix, Arizona, 2016. Retrieved May 2020 from https://www.rss.org.bb/wp-content/uploads/2018/03/Executive-summary-RSS-TYG-Report-.pdf

Kraus, Molly, Kristine Chan, Alfonso Martin, Loraine Park, Jorja Leap, Laura Rivas, Kim Manos, Karen Hennigan and Kathy Kolnick. *GRYD Gang Prevention 2017 Report*. 2017. Retrieved May 2020 from https://www.lagryd.org/sites/default/files/reports/GRYD%20Prevention%20Report_Final.pdfKrumpal, I. "Determinants of social desirability bias in sensitive surveys: a literature review." *Qual Quant* 47, 2013. Retrieved July 2020 from https://link.springer.com/article/10.1007/s11135-011-9640-9?con=1084525252525253FelqCampaignId52172525252525252525253Fcon10845

Littell, Julia, Melania Popa, and Burnee Forsythe. "Multisystemic therapy for social, emotional, and behavioral problems in youth aged 10-17." *Cochrane Database of Systematic Reviews* 4: 1–51. 2005.

Miguel, Edward and Michael Kremer. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica, Vol. 72, No. 1,* 2004. Retrieved May 2020 from http://cega.berkeley.edu/assets/cega_research_projects/1/Identifying-Impacts-on-Education-and-Health-in-the-Presence-of-Treatment-Externalities.pdf

Murray, Michael, A.V. Swan, S. Kiryluk, and G.C. Clarke. "The Hawthorne Effect in the Measurement of Adolescent Smoking." *Journal of Epidemiology and Community Health.* 1988. Retrieved May 2020 from https://jech.bmj.com/content/jech/42/3/304.full.pdf

Scott, J. "Children as respondents: The challenge for quantitative methods." *Research with children: Perspectives and practices, 98-119.* New York: Psychology Press. 2000.

Sexton, Thomas and Charles Turner. "The Effectiveness of Functional Family Therapy for Youth with Behavioral Problems in a Community Practice Setting." *Journal of Family Psychology, Vol. 24, No. 3, 339-348*. 2010Sutton, Heather and Inder Rupran. *Restoring Paradise in the Caribbean: Combatting Violence with Numbers.* Washington, DC: Inter-American Development Bank, 2017. Retrieved May 2020 from https://publications.iadb.org/publications/english/document/Restoring-Paradise-in-the-Caribbean-Combatting-Violence-with-Numbers-Executive-Summary.pdf

Thornberry, Terence, Brook Kearley, Denise Gottfredson, Molly Slothower, Deanna Devlin, and Jamie Fader. "Reducing Crime Among Youth at Risk for Gang Involvement: A Randomized Trial." *American Society of Criminology: Criminology & Public Policy, Vol. 17, Iss: 4.* 2018. Retrieved May 2020 from https://doi.org/10.1111/1745-9133.12395

UNICEF. Innocenti Research Briefs on Adolescent Wellbeing. 2017. Retrieved May 2020 from https://www.unicef-irc.org/adolescent-research-methods/

Wolf, A.M., A. Del Prado Lippman, C. Glessmann, and E. Castro. *Process evaluation for the Office of Neighborhood Safety.* Oakland, CA: National Council on Crime and Delinquency, 2015. Retrieved May 2020 from https://www.nccdglobal.org/sites/default/files/publication_pdf/ons-process-evaluation.pdf

World Health Organization. Preventing Youth Violence: An Overview of the Evidence. 2015.

U.S. Agency for International Development
1300 Pennsylvania Avenue NW
Washington, DC 20523