NBER WORKING PAPER SERIES

VIOLENT DISCIPLINE AND PARENTAL BEHAVIOR: SHORT- AND MEDIUM-TERM EFFECTS OF VIRTUAL PARENTING SUPPORT TO CAREGIVERS

Lelys I. Dinarte Diaz
Saravana Ravindran
Manisha Shah
Shawn M. Powers
Helen Baker-Henningham

Working Paper 31338
http://www.nber.org/papers/w31338

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023

We are thankful to Anna Aizer, S Anukriti, Kathleen Beegle, Janet Currie, Andrew Dustan, Bilge Erten, Emanuela Galasso, Melanie Guldi, and Jevay Grooms for their useful comments and feedback. We also appreciate comments from participants at the 2023 NBER Childrens Meeting, 2023 EDW, 2023 WECJr, PacDev 2023, 1st WAP Workshop, 2023 LACEA IEN Meeting, Atlanta Workshop on Public Policy and Child Well-Being 2023, Development Research Group (World Bank), IFPRI, Exeter, University of Western Australia, NUS, PUC Chile, Universidad Los Andes, and Universidad Andres Bello. Carlos Guzman, Gabriela Lecaro, and Nathy Andrade provided superb research assistance. We are truly grateful for the support of the Early Childhood Commission (ECC) for implementing the intervention. We thank Taja Francis for supporting the development of the virtual Irie Homes Toolbox and the training of officers. This work was supported by the World Bank SIEF and ELP Trust Funds and the Research Support Budget. Ravindran is funded by a startup grant at the Lee Kuan Yew School of Public Policy, National University of Singapore. This research project's protocol was reviewed and approved by the UWI Institutional Review Board in Jamaica (approval ID No. CREC-MN.86, 20-21) and by the IPA Research Ethics Committee (approval ID No. 15586). The trial was registered at the American Economic Association RCT registry (AEARCTR-0008266). The authors have no conflicts of interest to report. The findings, interpretations, and conclusions expressed in this report are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, its Executive Directors, the governments they represent, or the National Bureau of Economic Research.

Violent Discipline and Parental Behavior: Short- and Medium-term Effects of Virtual Parenting Support to Caregivers

Lelys I. Dinarte Diaz, Saravana Ravindran, Manisha Shah, Shawn M. Powers, and Helen Baker-Henningham

NBER Working Paper No. 31338
June 2023
JEL No. I25,J12,J13,J16

ABSTRACT

Approximately 75% of children aged 2 to 4 worldwide are regularly subjected to violent discipline across the globe. We study the impact of a virtually-delivered intervention on positive parenting practices in Jamaica. We find the intervention improves caregiver knowledge (0.52 SD) and attitudes around violence (0.2 SD) and leads to meaningful changes in caregiver disciplining behaviors, with a 0.12 SD reduction in violence against children. Treatment children also experience fewer emotional problems (0.17 SD). When we return nine months later, we also find reductions in caregiver depression (0.12 SD), anxiety (0.16 SD), and parental stress (0.16 SD) for treatment caregivers. The virtual delivery has important scalable policy implications which could help decrease violence against children across the globe.

Lelys I. Dinarte Diaz
The World Bank
Development Economics Research Group
1818 H Street NW
Washington, DC 20433
and CESifo
ldinartediaz@worldbank.org

Saravana Ravindran
Lee Kuan Yew School of Public Policy
National University of Singapore
469B Bukit Timah Road
Level 2 #02-08 Li Ka Shing Building
Singapore 259771
Singapore
saravana@nus.edu.sg

Manisha Shah
Department of Public Policy
University of California, Los Angeles
Luskin School of Public Affairs
3250 Public Affairs Building
Los Angeles, CA 90095-1656
and NBER
manishashah@ucla.edu

Shawn M. Powers
The World Bank Group
1818 H St. NW
Washington, DC 20433
spowers1@worldbank.org

Helen Baker-Henningham
Bangor University
School of Psychology
Brigantia Building
Bangor, Gwynedd
Wales, LL57 2AS
United Kingdom
and University of the West Indies
h.henningham@bangor.ac.uk

A data appendix is available at http://www.nber.org/data-appendix/w31338
A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/8266

1 Introduction

Approximately 75% of children aged 2 to 4 worldwide—close to 300 million—are regularly subjected to violent discipline (physical punishment and/or psychological aggression) by parents and/or caregivers at home (UNICEF, 2017).[1] Such exposure to violence can hinder children's development and undermine their sense of self-worth (Boden et al., 2007; Fry et al., 2018; Mersky and Topitzes, 2010). Moreover, research shows children who are victims of abuse and neglect are more likely to exhibit risky behaviors as teenagers (Hamby et al., 2011); be absent from school, and have higher levels of aggression, mental distress, and social problems (Lansford et al., 2002). These children can negatively affect peers' test scores and behavior in the classroom (Carrell and Hoekstra, 2010). Early exposure to violence also negatively affects outcomes during adulthood (Doyle Jr and Aizer, 2018), resulting in worse labor market outcomes (Currie and Spatz Widom, 2010) and/or increased involvement in crime (Currie and Tekin, 2012; Sviatschi, 2022).

In this paper we test whether a virtually delivered, scalable intervention on positive parenting practices can improve caregivers attitudes and behaviors related to violence against children in Jamaica, where 82% of children aged 2-14 are regularly subjected to violent discipline at home (UNICEF, 2022). Within Latin America, Jamaica ranks third highest in terms of violence against children (Haiti and Suriname rank first and second) (UNICEF, 2022). We adapt an in-person Jamaican behavior change violence-prevention parenting program—the Irie Homes Toolbox (IHT)—for virtual delivery. The intervention is targeted at parents of 2-to-6-year-old children (we refer to the eldest child aged 2 to 6 in each household as the "target child"). The virtual IHT is a 10-week program that includes content related to four core concepts: i) building positive relationships between parent and child, ii) preventing misbehavior, iii) managing misbehavior, and iv) supporting children's emotional self-regulation (Francis and Baker-Henningham, 2020). The content is delivered through three components: First, participants receive three SMS messages per

[1] This exposure to violence can start even earlier. According to data from 30 countries, nearly half of children aged 12 to 23 months are subjected to corporal punishment and a similar proportion are exposed to verbal abuse at home (UNICEF, 2017).

week (for a total of 30 SMS messages). Each SMS message briefly describes the techniques caregivers should practice from each of the four core elements of the intervention. Second, caregivers receive access to a data-free app with weekly content (videos and other related materials). Third, participants were offered the opportunity to join weekly, one-hour, virtual group parenting sessions led by field officers of the Early Childhood Commission in Jamaica (ECC).

To measure the impacts of the virtual program, we collected three rounds of data: before the start of the intervention (baseline), right after its completion (short-term follow-up), and 9 months later (medium-term follow-up). We collected data on caregivers' attitudes and perpetration of violence against children, caregivers' mental distress, parental self-efficacy, and changes in caregiver support networks. We also collected data on conduct and emotional problems of their children as well as demographic characteristics of other household members. We complement this household survey data with administrative data from TrendMedia on SMS messages sent and App usage, reports on attendance at the virtual sessions provided by ECC field officers, and focus group discussions with caregivers.

Intent-to-treat estimates show that the intervention led to significant improvements in caregivers' attitudes toward violence against children (VAC). Caregivers in the treatment group improved attitudes by 0.2 SD in the VAC index ($p < 0.01$) in the short run. This is driven by a 0.19 SD improvement in the physical VAC sub-index ($p < 0.01$) and a 0.12 SD improvement in the psychological VAC sub-index ($p < 0.05$). These effects are persistent, with caregivers in the treatment group improving attitudes by 0.14 SD on the VAC index ($p < 0.05$) in the medium-term follow-up.

Consistent with improvements in attitudes toward VAC, caregivers in the treatment group also changed their child disciplining behaviors. Caregivers in the treatment group reduced violence against the target child (VATC) by 0.12 SD ($p < 0.05$). We construct sub-indices for the physical and psychological VATC and find 0.14 SD ($p < 0.05$) and 0.1 SD ($p < 0.1$) reductions in these indices in the short term. These effects persist in the medium-term, where we estimate a 0.13 SD reduction in VATC ($p < 0.05$). Furthermore,

while we do not observe statistically significant impacts on caregivers' mental well-being at the first follow-up, we show reductions in caregiver depression (0.12 SD, $p < 0.1$), anxiety (0.16 SD, $p < 0.05$), and parental stress (0.16 SD, $p < 0.05$) at the second follow-up. Consistent with improvements in caregiver attitudes and disciplining practices, we find that target children in the treatment group scored 0.17 SD lower on the index of the emotional problem in the short term ($p < 0.01$).

We introduce a simple conceptual framework to understand the results and potential mechanisms. We hypothesize that the intervention may have improved caregivers (i) knowledge about positive parenting practices, (ii) self-efficacy, and/or (iii) support networks. We show that caregivers in the treatment group learned from the material and scored 0.53 SD higher ($p < 0.01$) on the information module relative to caregivers in the control group at the first follow-up. The treatment effects on caregiver knowledge are persistent: the treatment group scored 0.39 SD higher ($p < 0.01$) on the information index relative to caregivers in the control group in the second follow-up. Administrative and survey data show that 92% of weekly messages were sent and successfully received by treated caregivers. Ninety-seven percent of caregivers reported they read the messages, and 98% of the caregivers who read the messages found them to be useful.

We also show positive impacts of treatment on caregivers' self-efficacy at the second follow-up (0.21 SD, $p < 0.01$). However, it appears that caregiver parenting support and/or borrowing networks are not relevant in explaining the improvements. The positive treatment effects and analysis of mechanisms suggest that the intervention provided caregivers with the necessary parenting tools and boosted their confidence in their own parenting skills.

We examine the robustness of the results in the following ways. First, we address potential concerns related to experimenter demand effects since many of the outcomes are self-reports. Following Asadullah et al. (2021) and Dhar et al. (2022), we test whether the treatment had an effect on the caregivers' social desirability index.[2] We do not find

[2]This index captures the study participant's individual-level propensity to misreport sensitive items, which indicates whether or not the respondent is driven by the need for social approval.

evidence that desirability bias changed among treated caregivers relative to those in the comparison group. In additional robustness checks, we include the social desirability index as a control variable and this does not impact the ITT estimates or standard errors. Second, we examine whether the intervention displaced violence from the target child to other children in the household. We find the opposite. Including data on the eldest child aged 7-12 in the household, we find 0.14 SD and 0.15 SD reductions in the index for violence against these children in the short- and medium-term ($p < 0.01$). Third, we explore issues related to selective attrition and show that our results are robust to attrition. Fourth, we verify the robustness of our results to the inclusion of additional control variables as selected by a double LASSO algorithm. Our results remain similar in terms of magnitude and inference. Lastly, we take a more agnostic approach to the structure of the standard errors and estimate them using a randomization inference approach. We find that the magnitude of the *p*-values is similar to the magnitudes of *p*-values obtained by estimating heteroskedasticity-robust standard errors.

Our paper makes several important contributions to the literature. First, as far as we know, this is the first causal study of a virtually-delivered intervention to caregivers with the primary goal of reducing violence against children. The majority of previous early childhood interventions are more costly in-person interventions that focus on nutrition as well as cognitive and socio-emotional development of children (Carneiro et al., 2023; Bos et al., 2022; Mehrin et al., 2022; Chandra et al., 2021; Heckman et al., 2020; Attanasio et al., 2020; Hamadani et al., 2019; Levere et al., 2016; Attanasio et al., 2014; Macours et al., 2012; Paxson and Schady, 2010) and few in-person studies target violence against children as the primary outcome (de Simone et al., 2022; Francis and Baker-Henningham, 2021; Lachman et al., 2020; Cluver et al., 2018; Altafim and Linhares, 2016; Knerr et al., 2013).

Second, we contribute to a small and nascent literature on digital and low-cost parenting programs, since most parenting programs occur in person. The few interventions that have been virtually-delivered (see, for example, Smith et al. (2023); Amaral et al. (2021); Barrera et al. (2020); Widen et al. (2020); York et al. (2019); Cortes et al. (2019); Doss et al. (2019); Hurwitz et al. (2015)) do not target violence against children as the main outcome. In low-income settings, the widespread availability of mobile phones, the high incidence

of violence against children, and the general social acceptance of this issue make our intervention potentially attractive, but the effectiveness of this type of digital programming is not yet well understood. Lastly, this study contributes to questions of scalability by testing an alternative and low-cost delivery mode of a parenting intervention. Overall, the high prevalence of child maltreatment and its negative long-term impacts call for innovative and effective strategies to address this global problem.

2 Intervention and Conceptual Framework

2.1 Intervention

The Irie Homes Toolbox (IHT) consists of a behavior change violence prevention program targeting parents of 2-to-6-year-old children. The content of the program is based on evidenced-informed parenting practices that improve parenting behavior and reduce child behavior problems (Chorpita and Daleiden, 2009; Garland et al., 2008; Michie et al., 2011). We generated a digitally adapted version of the Irie Homes Toolbox (vIHT) for this study.

The vIHT intervention includes content related to four key concepts: i) building positive relationships between parent and child (e.g. praise, child-led play, involving the child in everyday activities), ii) preventing misbehavior (e.g. understanding why children misbehave, giving children independence and autonomy, giving clear instructions, setting rules and expectations, modeling appropriate behavior), iii) managing misbehavior (e.g. redirecting child, withdrawing attention, setting limits, giving appropriate consequences) and iv) child emotional regulation and stress reduction techniques (Francis and Baker-Henningham, 2020).[3]

The vIHT content was delivered over a 10-week period via smartphones. First, ben-

[3]The in person version of the program consists of an eight-week, group-based, parenting intervention, delivered through community preschools. Francis and Baker-Henningham (2021) show that the in person program reduced parents' use of harsh punishment by 0.29 SD, increased caregiver involvement with their child by 0.30 SD, and reduced child behavior difficulties among children at or above the 50th percentile on initial behavior difficulties by 0.36 SD.

eficiaries received three SMS messages per week (30 SMS messages in total) relating to content from the Irie Homes Toolbox.[4] Each SMS briefly describes the techniques parents should practice during the week. It also includes a link to the content embedded in the program App and information on the Irie Challenge for the week. The Irie Challenge consists of suggestions on how to put the strategies learned from the program into practice with children. Second, caregivers received access to a data-free App. This program App included videos of parents utilizing the strategies with their children, the Irie Tower (which depicts the list of all strategies thought during the program), and the Irie Challenge. The content was uploaded and updated every week during the 10 weeks.[5]

Lastly, caregivers were offered the opportunity to join a virtual parent group that met once per week for ten weeks through GoogleMeet video calls to discuss the specific topic of the week with an early childhood education specialist. We partnered with the Early Childhood Commission (ECC) in Jamaica for the implementation of this third component. ECC officers were trained in the curriculum and its implementation by the Irie Toolbox Team based in the Caribbean Institute for Health Research. The groups included 8 to 9 participants and were formed randomly (see more details in the research design section). Participant caregivers received data packages that allowed them to join the virtual groups. During these sessions, the messages received by text and video were reinforced through discussion and practice. Moreover, ECC officers sent e-summaries of the session via WhatsApp at the end of the session.

2.2 Conceptual Framework

Our analysis is organized around the conceptual framework presented in Figure 1 which was also pre-registered at the AEA. This framework approaches the incipient problem of high levels of violence against children in households in low- and middle-income countries. We expect the intervention to potentially impact at least three potential primary out-

[4]See the content delivered during each weekly session in Table A1 in the appendix. The complete list of SMS messages can be found in Table A2 in the Appendix.

[5]See Figure A1 in the Appendix with examples of materials corresponding to Week 4.

comes. First, caregivers could learn about positive parenting practices and violence prevention. This could then impact their attitudes and behaviors related to violence against children. If we observe changes on these margins, then secondary outcomes of interest will be related to caregiver mental health and/or children's emotional/behavioral problems (Chorpita and Daleiden, 2009; Garland et al., 2008; Francis and Baker-Henningham, 2021).

In terms of potential change pathways or mechanisms besides increased parental knowledge on violence prevention parenting, parental self-efficacy, and/or a parent's belief in their ability to perform the role of parent successfully could be important mechanisms (Wittkowski et al., 2017). Evidence shows that caregiver's perceived support is associated with parental self-efficacy (Fang et al., 2021). Since the intervention aims to help caregivers through the provision of information and skills training on non-violent practices, this sense of support could assist them in increasing their parental self-efficacy. Moreover, considering the strong association between parental stress, maternal depression, and parental self-efficacy (Fang et al., 2021), the intervention could indirectly improve caregiver's mental health.

Lastly, social networks and the support that network members provide could be essential resources for caregivers in sustaining their caregiving role. By joining the virtual meetings, caregivers could expand their support networks through their interactions with other caregivers. For example, they could share information, knowledge, and experience about improving attitudes and violent behaviors toward their children. Information transmission has been shown to be a mechanism operating in programs that exploit variation in peers and networks (Dahl et al., 2014). Moreover, an improvement in caregivers' networks can also indirectly improve their mental health. As evidence shows, caregivers' perceptions of support networks, including family and friends, have been linked to their health status (Owen and Anderson, 2017; Balaji et al., 2007).

3 Experimental Design and Data

3.1 Participant Recruitment and Enrollment

We mainly recruited participants via SMS messages to Digicel's customers. Digicel is a mobile phone network provider headquartered in Jamaica, and is the main provider of mobile phone services in the country. As shown in Table A3, 93.4% of participants were recruited via this channel. We recruited the rest of the participants through the ECC and preschool principals and/or via social media. We partnered with TrendMedia,[6] a subsidiary of Digicel Group, to send every participant a link to an enrollment survey through these three channels. The enrollment survey included the following eligibility criteria: caregivers had to 1) live in the same house with at least one child aged 2 to 6 years old, 2) have access to a smartphone or tablet, and 3) provide consent to participate in the intervention and study. We also asked about gender and parish of residence[7] in the enrollment survey, but these two variables were not part of the eligibility criteria.[8] A potential concern with access to a smartphone or tablet as eligibility criteria is that it might shrink the potential sample of participants. However, mobile subscription is large in Jamaica. Data from the Office of Utilities Regulation indicates that mobile coverage in Jamaica was close to 100% in 2021 and 70.3% of the population had access to smartphones (Operator Watch, 2021).[9]

Digicel's SMS messages did not target particular parishes. Figure A5 in the Appendix shows a parish-wise comparison of the distribution of our sample relative to the distribution of the population of caregivers in Jamaica. Despite a few exceptions in Kingston and Saint Andrew and Saint Catherine, the proportions are comparable.

As Figure 2 shows, we identified a total of 1,993 eligible caregivers based on the criteria

[6] TrendMedia offers business solutions based on information and communication technologies, to corporate clients, SMEs, and governments.

[7] Parishes are the primary unit of local government in Jamaica.

[8] See SMS messages Snapshot in Figure A2 and snapshots of the full survey in Figures A3 and A4.

[9] See more details in https://our.org.jm/sectors/telecommunications/telecommunications-market-information-data/

mentioned above. In August 2021, we contacted all the 1,993 caregivers and collected baseline data from 1,113 individuals distributed across 14 parishes. The remaining 880 individuals did not complete the baseline survey for several reasons, including they did not provide a correct phone number (14%); we were directed to voicemail (32%); they changed their mind and decided not to participate in the study (24%); etc.

3.2 Randomization

We randomly assigned the sample of 1,113 enrolled caregivers to either the treatment or the control group with equal probability. Our sample includes 557 caregivers in the treatment group and 556 caregivers in the control group. Caregivers in the control group also received three SMS messages per week (30 SMS messages in total) with content related to good practices to avoid COVID-19.[10] Randomization was stratified by gender of the caregiver (male or female) and the mode of recruitment into the study (SMS messages campaign or ECC/Principal referral and social media campaign). Appendix Table A3 indicates the size of each stratum.

3.3 Data Collection

We collected data at baseline (before the intervention) in August 2021 and conducted two follow-up rounds (one right after the intervention ended and another one 9 months later). We administered phone-based surveys in all data collection rounds. We piloted the survey instrument before conducting the baseline (between May and June 2021). The intervention was completed at the end of November 2021, and we collected the first round of follow-up data in December 2021 from 985 caregivers (an 88.5% response rate). We timed the first follow-up survey to test the short-term effects of the program and to minimize attrition. To measure the medium-term effects of the vIHT, we collected a second round of follow-up data in August-September 2022. We were able to collect data from 705 caregivers (63.3% response rate). All data collected during the three rounds are self-reported. Participants

[10] See the complete list of SMS messages sent to the control group in Table A4 in the Appendix.

received a small monetary incentive (US$2.50) to complete each of the three surveys. We also conducted focus group discussions with 43 study participants in April and May 2023. A detailed description of this qualitative study is presented in Appendix A2.

A. SMS viewership, attendance, App use, and learning

We collected administrative data on SMS delivery and whether caregivers logged in to the App and the time (in minutes) they used the App from TrendMedia. ECC officers collected attendance data for the virtual groups for each caregiver. In addition, we collected self-reports from individuals on whether they read the text messages, how useful they found them, etc. We also asked questions to measure whether caregivers learned some concepts and practices that were taught in the program in the two follow-up rounds to measure learning from the information provided.

B. Main outcomes[11]

Attitudes toward violence against children: We use an adapted version of the UNICEF Multiple Indicator Cluster Survey (MICS) questionnaire (UNICEF, 2011) to measure parental attitudes toward physical and psychological violence against children at baseline and follow-up rounds. The adapted instrument includes 5 items asking about some attitudes such as if they agree that a good parent can slap the child if he misbehaves, if shouting and yelling would make the child more obedient, etc. Due to time constraints, we selected 5 out of 13 items with the greatest variation based on the results from the survey instrument pilot that was conducted prior to baseline data collection.

Violence against children: We use the UNICEF MICS questionnaire to measure caregivers' perpetration of physical or psychological violence against children. The adapted instrument includes 5 items asking about physical violence (hitting the child with a bare hand or with an object) and/or psychological violence (shouting, yelling, or screaming at the child; saying to send the child away; threatening to hit the child). We asked each caregiver about perpetrating these violent acts to the "target child" or to another older child within

[11] All outcomes were registered at the American Economic Association RCT registry - AEARCTR-0008266. A detailed description of the items included in the estimation of the outcomes and other variables is presented in Appendix A1.

the household (eldest child between 7 to 12 years old). Using these reports, we created two indices: violence against the target child and violence against any child within the household.

C. Secondary Outcomes

Caregiver's mental health: We administered the Patient Health Questionnaire (PHQ-2 survey, Kroenke et al. (2003)) and a question on having difficulty sleeping at night. We measure anxiety using the Generalized Anxiety Disorder (GAD-2, Donker et al. (2011)) instrument. We included these instruments during all three rounds of data collection. Moreover, during the second follow-up, we also included questions from the 18-items Parental Stress Scale (PSS-18, Berry and Jones (1995)). Our main outcomes of interest are indices of depression, anxiety, and stress.

Child conduct and emotional problems: We use 10 items related to children's conduct and emotional problems (5 items each) from the Strengths and Difficulties Questionnaire (SDQ) instrument to measure a child's behavior. We collected this information in each of the survey rounds. Each question is answered on a 0-2 scale (Not true, somewhat true, certainly true).

D. Potential Mechanisms

Parental self-efficacy: We measure parental self-efficacy at baseline and during the first follow-up round using the 5 items from the Brief Parental Self Efficacy Scale (BPSES) instrument. The scale asks parents about their agreement with statements that can describe their ability to improve a child's behavior. For the second follow-up, we adapted the Tool to Measure Parenting Self-Efficacy (TOPSE) for more detailed questions relating to discipline and self-acceptance.

Caregiver's support networks: The effectiveness of positive parenting programs can be driven by support networks for participant caregivers. To test this potential mechanism, we collect information on whether caregivers obtained support from friends, family, or professionals to solve parenting or financial issues. We asked how many people they could reach out to in case they need to talk about issues related to parenting and child rearing or borrowing money.

3.4 Baseline Summary Statistics

Table 1 reports summary statistics of the sample. On average, control group caregivers are 33 years old, have 14 years of education, and 85% of them are female. Furthermore, 38% of the caregivers are married, 79% reported being employed and the average income in the past month was USD 910. With respect to household characteristics, the average household has 4.6 members, with approximately 2 of them being children under 17. Target children are, on average, 4.1 years old and are gender-balanced (48% are female).

Caregivers exhibit relatively high support for violence against children, with 37% agreeing to the statement: "Shouting, yelling, and threatening to slap will not harm the child." Furthermore, the average caregiver draws on harsh conduct to discipline their children 2 to 6 years old approximately 1.5 days per week ("Shouted, yelled, or screamed at him/her"). For child behavior-related outcomes, we observe that 41.7% of children exhibit conduct problems, while 24.7% display emotional problems. Average prevalence rates of depression and generalized anxiety disorder are 20% and 13%, respectively.

How do caregivers in our sample compare to the representative caregiver of a child aged 2-6 in Jamaica? We compare our sample of 1,113 caregivers with the Jamaica Survey of Living Conditions (JSLC) 2019. The JLSC is a living standards measurement survey that is representative of the Jamaican population. We restrict the JSLC sample to only caregivers with a child aged 2-6 and compare key demographic variables for which data is available across surveys in Appendix Table A6. The last column of the table provides p-values for the comparison of means across the two samples. We find that children across the samples are quite similar in terms of age and gender. However caregivers in our sample are slightly younger, have one more year of education and are more likely to be married and employed. Caregivers in our sample seem to have slightly better socioeconomic status than the average caregiver in Jamaica.

We also address the potential external validity concern that our results may be specific to the COVID-19 pandemic period during which the intervention was delivered in Jamaica. To better understand counterfactual rates of violent disciplining by caregivers and mental health absent the intervention, we present summary statistics of these vari-

ables for control group caregivers for each of the three rounds of surveys in Table A7. The rates of violent disciplining and responses to questions regarding depression and anxiety are very similar across baseline, first, and second follow-up. Importantly, this time period spans August 2021 to September 2022, thereby covering a significant period after the relaxation of all COVID restrictions in Jamaica in March 2022. Table A7 shows that the counterfactual rates of violent discipline and mental health are stable across the various time periods, and we do not observe any meaningful changes in counterfactual trends post-COVID restrictions.

3.5 Baseline Balance

We test for balance in pre-intervention outcomes and socio-demographics between treatment and control groups. These results are presented in Table 1. Columns 2 and 5 present sample means for each variable (control and treatment) and columns 3 and 6 present the standard deviations (SD). Column 7 provides *p*-values for t-tests for equality of means between the treatment and control groups. With the exception of two out of 27 variables tested, we do not find significant differences in these variables across treatment and control groups at *p*-values less than 0.1. The only two differences we find are that caregivers in the treatment group told their children they would send them away less in the past seven days (0.14 days in the treatment group versus 0.23 days in the control group; $p = 0.023$), and were more likely to suffer from generalized anxiety disorder at baseline (17.1% in the treatment group versus 13.3% in the control group; $p = 0.084$). The *p*-value for the overall F-test of joint orthogonality is 0.968, highlighting that jointly, the means of the variables are not statistically distinguishable across the treatment and control groups. The randomization produced comparable treatment and control groups.

4 Estimation Framework

4.1 Empirical methods

To study intent-to-treat (ITT) impacts, we estimate Ordinary Least Squares (OLS) regressions using the following Analysis of Covariance (ANCOVA) specification for caregiver i in period t and stratum s:

$$Y_{i,t} = \beta_0 + \beta_1 T_i + \beta_2 Y_{i,t-1} + \gamma_s + \varepsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ refers to the outcome variable of interest of caregiver i as measured at first or second follow-up, defined in Section 3.3. T_i is an indicator variable capturing the assignment of caregiver i to the treatment group, and $Y_{i,t-1}$ refers to the outcome variable of interest measured at baseline. γ_s captures stratum fixed effects for the four strata; the interaction between the gender of the caregiver (female or male) and the two modes of recruitment (SMS messages or social media). The estimate of β_1 captures the ITT effect of the treatment. We estimate and report heteroskedasticity-robust standard errors. As a robustness check, we follow a more agnostic approach to the structure of the standard errors (or a potential fuzzy clustering) and estimate randomization inference standard errors. Randomization inference gives us precise *p*-values based on the empirical distribution of all estimated treatment effects that could arise under our design and data (after randomly reassigning the treatment status 1,000 times) under the null hypothesis of no effect for any unit.

To address potential concerns relating to multiple hypothesis testing, we construct indices for broad families of outcomes using Anderson (2008). Summary index tests offer three advantages: (i) they are robust to over-testing because each index represents a single test, (ii) they provide a statistical test for whether a program has a "general effect" on a set of outcomes, and (iii) they are potentially more powerful than individual-level tests by reducing random error in each outcome measure (Anderson, 2008). Each summary index is a weighted mean of several standardized outcomes, where the weights are calculated

to maximize the amount of information captured in the index using an efficient generalized least squares (GLS) estimator.[12] We orient caregivers and child outcomes such that a reduction in the index is always an improvement in the outcome of interest.

We also conduct Least Absolute Shrinkage and Selection Operator (LASSO) analysis to identify variables with strong relationships with $Y_{i,t}$, to assess their suitability for inclusion as controls in Equation (1) following Bruhn and McKenzie (2009). Since LASSO consistently only selected the outcome variable of interest measured at baseline ($Y_{i,t-1}$) for inclusion across all outcomes, we do not include other control variables in Equation (1).[13]

5 Short- and Medium-term Results

We use the Conceptual Framework and pre-analysis plan to guide the empirical work. We first test whether treatment was effective in improving parents knowledge of positive parenting practices related to violence against children. If treatment is effective, we might expect changes in caregiver attitudes and behaviors related to VAC.

We also investigate impacts on secondary outcomes related to caregiver mental health and child emotional problems (in Sections 5.4 and 5.5, respectively). Later we discuss the role of potential mechanisms.

5.1 ITT Impact on Violence Prevention Knowledge

To assess whether the vIHT content increased caregiver knowledge on positive parenting practices, we administered an information module "test" to all caregivers at endline. Table 2 presents ITT impacts of the treatment on caregiver knowledge. Columns 1 to 8 present the treatment impacts on eight statements relating to parenting practices. The eight statements were designed to assess understanding of the four key concepts of the intervention

[12] We show robustness of our results to the use of unweighted summary indices following Kling et al. (2007) in Section 5.

[13] Our main specification only includes the corresponding outcome variable of interest measured at baseline as a control, since it is selected consistently across all outcome variables. We assess robustness to the inclusion of other controls selected only for some outcomes in Section 7.

as outlined in Section 2 (two statements were asked for each of the four concepts). All statements are true; thus higher values represent greater knowledge. Panel A presents the short-term results from the first follow-up conducted immediately after the intervention ended. We observe that caregivers in the treatment group are significantly more likely to state that the statements are true relative to caregivers in the control group for six out of the eight statements (four statements significant at the 1% level, one statement significant at the 5% level, and one significant at the 10% level). To address potential concerns relating to multiple hypothesis testing, we aggregate the responses to the eight statements into one information index as outlined in Section 4. Column 9 of Table 2 shows that caregivers in the treatment group score 0.53 SD higher on the information module relative to caregivers in the control group. The impact is large in magnitude and statistically significant at the 1% level. Panel B presents the medium-term results from the second follow-up conducted nine months after the intervention ended. We see persistent treatment impacts on caregiver knowledge: the treatment group scored 0.39 SD higher on the information index relative to caregivers in the control group ($p < 0.01$). Table A8 (panel D) shows that the medium-term impact on the information index is not statistically different from the short-term impact.

5.2 ITT Impacts on Attitudes Toward Violence Against Children

Next we investigate the impact of treatment on caregiver attitudes toward violence against children (VAC). Figure 3 and column (1) of Table 3 (Panel A) show that caregivers in the treatment group improved their attitudes on VAC by 0.2 SD in the short-run (VAC index, $p < 0.01$) in the short-run. Panel A of Table A9 presents results from the first follow-up and breaks down the attitudes toward VAC into five individual components. The first three columns constitute the attitudes toward physical VAC, while columns 4 and 5 comprise attitudes toward psychological VAC. In the short term, caregivers in the treatment group are 3.9 percentage points less likely to agree with the statement that children need to be physically punished in order to bring up, raise, or educate a child properly (this is 32% reduction from the control mean). They are 6.2 percentage points (27%) less likely to agree

that a good parent slaps their child when they misbehave and 5.1 percentage points (53%) less likely to agree that when a child is beaten, he/she will stop doing the unwanted behavior ($p < 0.01$). The sub-index for attitudes toward physical VAC is shown in Figure 3 and column (2) of Table 3 (Panel A), and we estimate a 0.19 SD improvement in this sub-index ($p < 0.01$).

In terms of short-run changes in attitudes toward psychological VAC, caregivers in the treatment group were 2.3 percentage points less likely to agree that shouting and yelling makes the child more obedient and 2.7 percentage points more likely to disagree with the statement that "shouting, yelling, and threatening to slap will harm the child," although these results were not statistically significant at conventional levels. Figure 3 also presents the treatment impact on the sub-index for attitudes toward psychological VAC. As shown in column (3) of Table 3 (Panel A), at first follow-up, caregivers in the treatment group improve attitudes toward psychological VAC by 0.12 SD ($p < 0.05$).

These changes in attitudes persist into the medium-term. Figure 3 and column (1) of Table 3 (Panel B) show that caregivers in the treatment group improved attitudes toward violence against children by 0.14 SD ($p < 0.05$) in the medium-run. Analyzing the sub-indices for attitudes toward physical and psychological VAC, columns (2) and (3) of Table 3 (Panel B) show persistent impacts on attitudes toward physical VAC (0.15 SD reduction, $p < 0.05$), although the changes in attitudes toward psychological VAC are no longer statistically significant. Panels A and B of Table A8 shows that the medium-term impacts are not statistically different from the short-term impacts. Columns (1)-(3) of Table A10 show that the results are very similar when we use indices constructed following Kling et al. (2007).

5.3 ITT Impacts on Violent Behaviors Against Children

The results in Figure 4 and column (4) of Table 3 (Panel A) highlight that caregivers in the treatment group reduced violence against the target child (VATC) by 0.12 SD ($p < 0.05$). The figure also highlights the treatment impacts on the sub-indices for physical and psychological VATC. We estimate a 0.14 SD ($p < 0.05$) and 0.10 SD ($p < 0.1$) reduction in

these indices (columns 5-6 of Table 3, Panel A).[14] The larger treatment impact on physical versus psychological VATC is in line with the larger treatment effects for the attitudes toward physical versus psychological VAC. These results persist nine months after the end of the intervention: column (4) of Table 3 (Panel B) highlight that caregivers in the treatment group scored 0.13 SD lower at the second follow-up on the violence against target child (VATC) index ($p < 0.05$). Columns (5)-(6) of Table 3 (Panel B) show that we estimate statistically significant reductions in physical VATC (0.12 SD reduction, $p < 0.1$), but do not find statistically significant impacts on psychological VATC. Panels A and B of Table A8 show that the medium-term impacts are not statistically different from the short-term impacts. Columns (4)-(6) of Table A10 show that the results from the short and medium term are very similar when we use indices constructed following Kling et al. (2007).

Panel B of Table A9 presents the short-term impacts for the five components that comprise the VATC index. The first two columns are components of the physical VATC index, while columns 3, 4, and 5 comprise components of the psychological VATC index. Caregivers in the treatment group are 9.7 percentage points (25%) less likely to hit their child on the bottom, hand, arm, or leg with their bare hand ($p < 0.01$). In terms of psychological VATC, caregivers in the treatment group are 10.1 percentage points (14%) less likely to shout, yell, or scream at their child ($p < 0.01$) and 5.5 percentage points (8%) less likely to threaten to hit their child without doing so ($p < 0.1$). However, they are 4 percentage points (44%) more likely to report they tell their child they will send them away when misbehaving ($p < 0.05$). This is suggestive of caregivers in the treatment group substituting away from other more violent forms of discipline.

Overall, we find that the treatment led to positive and persistent impacts on caregivers' attitudes toward VAC, which in turn led to persistent reductions in VATC.

[14] The magnitudes of these effects are similar to the impacts from a digital stress management and positive parenting intervention in El Salvador on physical violence perpetrated by female caregivers (Amaral et al., 2021).

5.4 ITT Impacts on Caregiver Depression, Anxiety, & Stress

Harsh behaviors toward children may be explained by stress, anxiety, and frustration (Persson and Rossin-Slater, 2018; Bendini and Dinarte, 2020), which affect parental functioning through psychological well-being (Belsky, 1984; Belsky and Jaffee, 2006; Taraban and Shaw, 2018). We estimate ITT impacts on caregivers' well-being as measured by depression, anxiety, and parental stress. Depression and anxiety were measured at both follow-ups using the Patient Health Questionnaire (PHQ-2) and Generalized Anxiety Disorder (GAD-2) instruments, respectively. Parental stress was measured using the Parental Stress Scale (PSS-18) developed by Berry and Jones (1995) at the second follow-up only.

While we do not observe statistically significant impacts on caregivers' well-being at the first follow-up, Figure 5 and columns (1)-(3) of Table 4 (Panel B) show reductions in caregiver depression (0.12 SD, $p < 0.1$), anxiety (0.16 SD, $p < 0.05$), and parental stress (0.16 SD, $p < 0.05$) at the second follow-up. The lag in improvements in caregivers' mental health suggests that mental health improved only after they applied new tools learned from the intervention. Alternatively, caregivers may have needed time to apply their learning from the intervention.

5.5 ITT Impacts on Child Behaviors

The conceptual framework suggests that positive impacts on caregiver attitudes and behaviors can lead to positive impacts on child behaviors. Figure 6 and columns (4)-(5) of Table 4 present the ITT impacts of the intervention on the target child's conduct and emotional problems. As outlined in Section 3.3, conduct and emotional problems were measured using caregivers' responses to the Strengths and Difficulties Questionnaire (SDQ) instrument.

The estimates show that children in the treatment group scored 0.17 SD lower on the emotional problems index in the short term ($p < 0.01$). These children also scored 0.03 SD lower on the conduct problems index in the short term; however, this result is not statistically significant at conventional levels. In the medium term, though coefficients

are negative, we do not find statistically significant reductions in conduct and emotional problems. Panel C of Table A8 shows that the medium-term impacts are not statistically different from the short-term impacts.

5.6 Intervention Take-up

As highlighted in Section 2, the intervention consisted of three components: three SMS messages per week, access to a data-free App with vIHT content, and weekly virtual sessions with ECC officers. We cannot causally unpack the relative contribution of each component. However, in this section we explore take-up of the various components using both administrative data and self-reports in Table 5.

Panel A of Table 5 shows that 92% of the 30 SMS messages were sent by TrendMedia to caregivers. Panel B highlights that 91% of caregivers in the treatment group reported receiving SMS and/or WhatsApp messages as part of the intervention. Of the 499 caregivers who reported receiving the SMS and/or WhatsApp messages, 97% reported having read them. Furthermore, 98% of the 444 caregivers who read the messages found them to be useful. Using streaming data from the App, we also tracked the total duration of time spent accessing content on the App in Panel C. The mean duration spent on the App (across the 10-week intervention) was 6.9 minutes across all treatment groups. Moreover, the mean number of sessions accessed on the App was 1 out of 10. Results from the qualitative study show that the App was not as good a resource to deliver this intervention and that some adjustments will be required to make it more attractive/accessible to this population (Szekely, 2023). In Panel D, we show that across all caregivers assigned to the treatment group, the mean number of sessions attended was 4.6 (out of 10), and 79% of caregivers attended at least one session.[15] From the focus group discussions, we document that there was some substitution in take-up across the virtual meetings and SMS messages. When we asked participants with low virtual meeting attendance their reasons for not attending many sessions, one of the main reasons given was that it was because they knew

[15]This attendance rate is 23 percentage points lower than the rate for the in-person intervention evaluated in Francis and Baker-Henningham (2021).

they were going to receive similar materials via SMS/WhatsApp (Szekely, 2023).

Does attendance at a greater number of sessions lead to more learning? Acknowledging that selection into attendance at sessions is endogenous, we explore this dose-response relationship in Figure A6, where we present coefficient plots of the impact of session attendance on the caregiver information index. To improve precision, we group attendance into 5 pairs of possible combinations, i.e. attendance at one or two sessions, three or four sessions, and so on (where attendance at none of the sessions represents the base reference group). The regression is estimated over caregivers in the treatment group only. We observe an increasing dose-response relationship: while caregivers attending one to four sessions did not score significantly higher than caregivers who did not attend any sessions, caregivers attending five to ten sessions scored significantly higher on the information index relative to caregivers who did not attend any sessions. With the exception of attendance at seven or eight sessions, caregiver knowledge monotonically increases in the short term in the number of virtual sessions attended. Results from the second follow-up show a very similar pattern in the medium term, with caregivers attending seven to ten sessions scoring significantly higher on the information index relative to caregivers who did not attend any sessions. These dose-response results suggest that our results are likely driven by learning new information rather than the SMS messages being just nudges or reminders of practices that caregivers already knew.

In sum, although this intervention was delivered through three components, the take-up rate of the App was relatively low and attendance at the weekly groups was low, suggesting the SMS and/or WhatsApp messages might have played an important role.

6 Potential Mechanisms

As highlighted in Figure 1, our conceptual framework hypothesizes at least three potential change pathways: (i) information, (ii) self-efficacy, and (iii) support networks. These mechanisms are based on previous research. First, research has highlighted that a lack of knowledge and skills on parenting practices may lead to harsh and unhealthy parenting behaviors (Baker-Henningham and Francis, 2018). Second, responsiveness-oriented

parenting approaches have shown the importance of self-efficacy beliefs in parenting functioning strategies (Michl-Petzing et al., 2019). Responsiveness-oriented behaviors are based on acceptance and warm responses toward child actions (Landry et al., 2012) and parents' perceptions about their self-efficacy (i.e. the extent to which they consider themselves capable and prepared to raise a child and deal with the associated parenting tasks) has become a cornerstone for good parenting practices (Izzo et al., 2000; Hoover-Dempsey et al., 2005; Jones and Prinz, 2005). Third, prior work has highlighted that parenting beliefs, attitudes, and behaviors may also be influenced by parents' social networks (Cochran, 2019).

The results in Section 5.1 documented the impacts of the intervention on violence prevention knowledge (the first potential pathway). We find large increases in knowledge related to positive parenting practices. In Figure 7, we present ITT impacts on the second and third potential pathways - self-efficacy and support networks. For self-efficacy, we used the Brief Parental Self-Efficacy Scale (BPSES) at the first follow-up but adapted to the Tool to Measure Parenting Self-Efficacy (TOPSE) at the second follow-up for more detailed questions relating to discipline and self-acceptance. For networks, we collected day on parenting support and borrowing support networks.

Figure 7 shows positive impacts on caregivers' self-efficacy relating to self-acceptance at the second follow-up (0.21 SD, $p < 0.01$). We do not find statistically significant impacts on caregivers' parenting support or borrowing networks. The lack of significant impacts on support networks highlights that peer interactions during the weekly virtual sessions were likely limited. These findings are confirmed by the qualitative study. Participants across the different focus groups reported that they did not make any friends from the program and that they are not in touch with anyone from a virtual group. Some of the reasons they mentioned included the vast geographic spread of other participants (the sessions were not organized to take into consideration the proximity across participants) and the virtual nature of the meetings (Szekely, 2023).

The positive treatment effects and analysis of mechanisms suggest that the intervention provided caregivers with the necessary parenting tools and boosted their confidence

in their own parenting skills. Taken together, our results suggest that the harsh behaviors of the caregivers could be explained, to a large extent, by a lack of knowledge and skills, self-efficacy, and emotional self-regulation. As Baker-Henningham and Francis (2018) have suggested, integrating new interventions that aim to train parents in alternative discipline strategies could help to improve the quality of parenting and reduce violent behaviors against the children.

7 Robustness of the Results

In this section, we address potential concerns regarding the results presented in Section 5, such as experimenter demand effects, displacement of violence toward other children in the household, and differential attrition across treatment and control groups in the first follow-up. We also present the results of additional robustness checks.

7.1 Assessing Potential Bias Due To Experimenter Demand Effects

Self-reported measures are susceptible to potential experimenter demand effects when assessing sensitive information such as attitudes and perpetration of violence because participants' responses regarding sensitive topics might be influenced by social desirability bias (Aguero and Frisancho, 2021; Amaral et al., 2021). Recent empirical evidence suggests a limited quantitative importance of experimenter demand effects in some domains (Haaland et al., 2023; de Quidt et al., 2018). The concern is whether participants report statements on violent attitudes and behaviors differently from their attitudes and behaviors outside of the study environment and whether this is differential across the treatment and control groups.

We take the possibility of experimenter demand effects seriously and conduct several exercises to address this potential concern. First, we test the intervention's direct effect on a social desirability index (SDI) estimated using the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960) which we collected during the second follow-up. The Marlowe-Crowne Social Desirability Scale has been shown in different settings to be in-

formative about one's propensity to report in socially desirable ways when asked about physical and psychological violence perpetration and victimization. For example, Bell and Naugle (2007) and Fernández-González et al. (2013) show sizeable correlations between SDI and physical and psychological aggression.

First, we test whether treatment and control caregivers differ by SDI. Column (1) of Table 6 (Panel A) shows there are no statistically significant differences between treated and control caregivers on the SDI. Second, we include the SDI as an additional control variable in the main regressions. These results in columns (2)-(7) of Table 6 (Panel A) show that the estimated effects remain similar in magnitude and statistical significance.

Third, we are concerned that caregivers in the treatment group might have a higher propensity to give socially desirable answers, thereby biasing the estimated treatment effects. Our results in Panel B of Table 6 show that there are no heterogeneous treatment impacts by SDI; the interaction terms in the third row are not statistically significant. The only instance where the interaction term is statistically significant is in column (2) (overall attitudes toward VAC), driven by attitudes toward psychological VAC (column 4). However, the interaction term is in fact positive, and in the opposite direction of what we would expect from treated individuals who are inclined to report in socially desirable ways.

Lastly, we investigate the potential concern that in addition to behaviors relating to violence, the SDI may also capture social desirability relating to other non-violence related behaviors. To address this concern, we select 5 items out of the 13 items included in the original SDI instrument that are most closely related to violence and aggression (i.e., "I have deliberately said something that hurt someone's feelings") and create a new violence-focused SDI. We repeat the analysis on social desirability with this new scale, and the results are presented in Appendix Table A11. Overall, the main conclusions relating to social desirability are sustained.

Together, these findings suggest that it is unlikely that experimenter demand effects have biased respondents' likelihood to misreport their attitudes and behaviors about the use of violence with their children across the treatment and control groups.

7.2 Potential Displacement of Violence Within the Household

A potential concern might be that the intervention succeeded in reducing caregiver violent discipline against the target child, but this may have displaced violence toward other children in the household. To address this concern, we surveyed caregivers about violence against their eldest child aged 7-12 in addition to the target child. We stack the responses on caregiver attitudes and behaviors to violence against children to form a child-level dataset that allows us to study violence against the target child, as well as the eldest child aged 7-12.[16]

Figure 8 shows 0.14 SD and 0.15 SD ITT reductions in the index for violence against these children in the short- and medium-term, respectively ($p < 0.01$). We also find 0.1 SD ($p < 0.1$) and 0.13 SD ($p < 0.05$) reductions in the indices of physical violence in the short- and medium-term, respectively. Moreover, we also find that caregivers in the treatment group scored 0.13 SD ($p < 0.01$) and 0.14 SD ($p < 0.05$) lower in the indices of psychological violence against any of these children at the first and second follow-ups.

Overall, these results suggest that the intervention did not displace violence from the target child to other children in the household.

Another potential concern from the intervention is the displacement of violent disciplining from the caregiver participating in the program to the other caregivers within their household that did not join the program. To address this concern, we asked participants of the focus groups if they shared information with other caregivers within the household and if they observed changes in the dynamics with their partners. Results from the qualitative study suggest there were positive spillovers within the household. Participants reported that they shared information from the intervention with their partners. For example, they encouraged their partner to praise their children for good actions. Moreover, they reported that family dynamics had improved after the program. They reported trying to be more gentle with their children instead of immediately being aggressive and also

[16] Panel A of Table 1 shows that households in our sample have an average of 1.9 children aged 17 or below. Given the constraints of the phone survey, we collected information on one elder child in the household.

played games as a family (Szekely, 2023).

7.3 Assessing Potential Bias Due To Differential Attrition

Between the baseline and first follow up survey, we lost 128 individuals, so approximately 11.5% of the sample. We then lost another 280 individuals, or 25.2% of the sample between the short and medium term. We test for differential attrition between treatment and control groups as we are concerned attrition might bias the estimates. In particular, if the higher proportion of treatment group caregivers who attrited are more likely to exhibit violent disciplining behaviors, our estimates may overestimate the true treatment effects.

We assess differential attrition between treatment and control groups for each follow-up round and present the results of this analysis in Table 7. Column (1) shows that caregivers in the treatment group were 4.7 percentage points more likely to complete the first follow-up survey relative to caregivers in the control group. In column (2), we study whether the differential attrition in the first follow-up was correlated with any demographic characteristics or outcome variables measured at baseline. None of the interaction terms in this regression are statistically significant at the 10% level, suggesting no evidence of relationships between demographic characteristics or outcome variables and the differential attrition. In line with this result, we cannot reject the null hypothesis that all interaction terms between the treatment indicator and the relevant variables are not statistically significant (p-value of 0.528).

In the second follow-up, we find no significant differences in the probability of not completing the survey between treatment and control groups (Column 3). Moreover, we find that none of the relevant variables explain any differential attrition between treatment and control, we do not reject the null hypothesis that the interaction terms between the treatment indicator and all the relevant variables are not different from zero (p-value = 0.916).

To address potential differential attrition in the first follow-up, we estimate Lee bounds to account for sample selection (Lee, 2009) and present these results in Table A12. This procedure is a conservative estimate of the treatment effect, as it corresponds to extreme

assumptions about the missing information. We find that all upper and lower bounds significantly differ from zero except the upper bounds of the index of attitudes toward psychological VAC (p-value = 0.177) and violence against the target child (p-value = 0.177) (driven by the psychological violence dimension), suggesting that our results are overall robust to differential attrition.

As an additional check, we re-estimate the ITT impacts on caregiver attitudes and behaviors, as well as child outcomes using a balanced panel of caregivers that were present at both the first and second follow-ups. This addresses concerns of potential selection in who might be present at either follow-up round. The results are shown in Figures A7 - A10. Overall, the results are very similar to the estimates presented in Section 5.

7.4 Assessing Sensitivity From Selection of Control Variables

As we discuss in Section 4, we use a double LASSO approach to identify the variables that should be included in our estimations as controls. We find that LASSO consistently selects, across all outcomes, the measure of the outcome at baseline. LASSO also selects other variables that can be used as controls, but they vary across outcomes. For instance, for some outcomes age and gender are selected, whereas for others LASSO selects education level and household composition. Considering this, we test for the stability of our estimated coefficients after including the control variables selected by LASSO for each of our main outcomes. As we show in Table A13, the estimated coefficients and their statistical significance do not change after including these additional control variables selected by LASSO.

7.5 Estimating the exact *p*-values using randomization inference

To take a more agnostic approach to the structure of the standard errors, we estimate standard errors using the randomization inference (RI) approach. As discussed in Section 4, what RI allows us to do is to assign a p-value for a given treatment effect by observing where that treatment effect falls in the distribution of all possible estimated effects from

the 1,000 randomizations we simulate under the assumption of no effects (Blattman et al., 2021). As we show in the row "RI p-value" in Table A14 in the Appendix, the magnitude of p-value is similar to the magnitudes of p-values obtained by estimating heteroskedasticity-robust standard errors.

8 Heterogeneity

Recent work has highlighted that the treatment impacts of parenting interventions may differ for sub-groups of the sample. For example, Amaral et al. (2021) highlight differential effects of a digital stress management and positive parenting intervention by the gender of the caregiver, with male caregivers experiencing increased stress and anxiety while female caregivers saw no impacts on mental health. Baranov et al. (2020) study a psychotherapy intervention for prenatally depressed mothers in Pakistan, and show significant treatment heterogeneity along wealth for outcomes including parenting style and children's socio-emotional development. Francis and Baker-Henningham (2021) also show, in an evaluation of the in-person Irie Homes Toolbox in Jamaica, that there were significant reductions in behavior difficulties for children with above-median baseline levels of behavior difficulties. Studying such treatment heterogeneity is important for assessing the scale-up and policy applications of our study.

We investigate treatment impact heterogeneity in two stages. First, we use machine learning techniques to understand if there is any evidence of significant treatment heterogeneity for our primary outcomes of interest. Second, conditional on evidence of heterogeneity, we use traditional interaction term analysis to understand the dimensions along which this heterogeneity arises.

In the first stage, to understand if there is any evidence of heterogeneity, we use casual forest algorithms to estimate Conditional Average Treatment Effects (CATEs) following Athey et al. (2019), Athey and Wager (2019), and Chernozhukov et al. (2023). Given subject characteristics, we use a subject-specific treatment prioritization rule that assigns scores to subjects, with higher scores assigned to caregivers with the largest benefit from the treatment as given by the CATE. To quantify treatment benefits, we use the Targeting

Operator Characteristic (TOC), which compares the ATE in smaller groups defined by the prioritization rule to the overall ATE from treating everyone in the treatment group.

We present results investigating evidence of treatment heterogeneity for our primary outcome variables (information index, attitudes toward VAC index, and violence against the target child index) in Table 8. To do so, we categorize caregivers into high and low ATE groups by creating below and above median CATE sub-groups, and estimate the ATE in each group. We observe a statistically significant 0.56 SD difference in above versus below median CATE for attitudes toward VAC. This shows evidence of significant treatment heterogeneity for attitudes toward VAC. While the information index and violence against target child variables show large differences – 0.33 SD and 0.12 SD, respectively – these differences are not statistically significant. Therefore, we conclude that we do not have sufficient evidence suggesting significant treatment heterogeneity for the information index and the violence against target child index.

Figure 9 presents the TOC corresponding to attitudes toward VAC, where the x-axis denotes the top q-th fraction of individuals with the largest prioritization score. We pool our first and second rounds of follow-up data to estimate the CATEs and plot the TOC. We see significant treatment heterogeneity for attitudes toward VAC. For example, the top 20% of caregivers with the largest prioritization score show treatment effects that are approximately 0.4 SD larger than the ATE. The heterogeneity is statistically significant at the 5% level for most of the distribution, as indicated by the 95% confidence interval bars.

Given the evidence of significant heterogeneity for attitudes toward VAC, in the second stage of analysis, we investigate the specific dimensions along which this heterogeneity arises using a traditional interaction term analysis. These results are presented in Table 9. Motivated by the recent literature outlined at the start of this section, we consider heterogeneity along the following dimensions: (i) gender of the caregiver, (ii) income, (iii) baseline attitudes toward VAC, (iv) baseline conduct problems of the target child, and (v) baseline emotional problems of the target child. Panels A and B present the analysis for the short- and medium-term follow-ups, respectively.

We observe significant heterogeneity by baseline attitudes toward VAC: the treatment

effects are 0.19 SD and 0.21 SD larger in the short- and medium-term for caregivers with worse than median attitudes toward VAC at baseline (column 3). Columns (4) and (5) highlight that treatment effects are 0.2 and 0.18 SD larger for target children with above-median conduct and emotional problems, respectively. However, these results are only statistically significant in the short-run. We do not find any significant heterogeneity by gender of the caregiver or income.

Taken together, the results highlight significant heterogeneity in attitudes toward VAC, shown using machine learning techniques. Traditional interaction term analysis highlights the role of baseline attitudes toward VAC, as well as emotional and conduct problems of the target child in the short-run, in explaining some of this heterogeneity.

9 Cost Effectiveness

Table 10 shows the cost-effectiveness of our intervention relative to the face-to-face IHT in Jamaica (Francis and Baker-Henningham, 2021) and a cash transfer intervention in Mali (Heath et al., 2020). We show these comparisons because these interventions also reduced physical or psychological violence toward children or the female partner. Excluding the cost of developing the App, which is a fixed cost for the intervention component that was least used by caregivers, the vIHT cost USD 62.4 per person and reduced violent behaviors against children by 0.11 to 0.13 SD in the medium-term. The cost per 0.13 SD effect is between USD 62.4 and USD 73.75. In comparison, the face-to-face IHT cost USD 123.95 per person and reduced physical violence against children by 0.29 SD, yielding a USD 55.56 cost per 0.13 SD effect. The per person cost of the vIHT is a similar magnitude to the in person IHT. However, we note that the vIHT has greater potential to be scaled-up given the virtual delivery via smart phones. It is less demanding on staff capacity to conduct the virtual meetings relative to the number of professionals required to conduct the face-to-face activities. In contexts with a limited number of local professionals for early childhood development and parenting interventions, the implementation of face-to-face parenting interventions are more difficult to implement at scale.

In a cash transfer intervention in Mali, Heath et al. (2020) show that cash decreased

physical and psychological violence against the female partner (0.13 and 0.12 SD, respectively) and physical violence against children (0.17 SD). This translates to a per person cost between USD 496 and USD 702.8 per 0.13 SD effect size. While we note that cash transfer programs may aim to improve several other outcomes, this result highlights the strong cost-effectiveness of our intervention relative to cash transfers to reduce violent behaviors toward children.[17]

10 Conclusion

Violence against children is a global problem that will require new and creative solutions. We evaluate the impact of the virtual Irie Homes Toolbox, a positive parenting program for Jamaican parents of children aged two to six years. Our results show that the intervention improved caregivers' attitudes and violent behaviors against children in the short term. Furthermore, the program reduced children's emotional problems. Results in the medium term from a second follow-up indicate that these effects on caregivers' attitudes and violent behaviors against children persist nine months after the intervention ended. We also document important improvements in caregivers' mental health in the medium term. In terms of potential mechanisms driving these effects, we find evidence consistent with improvements in parental self-efficacy most likely due to the knowledge and skills they gained from the virtual Irie Homes Toolbox.

From a policy perspective, the positive effects of the intervention on attitudes and behaviors related to violent discipline against children provide evidence of the importance of developing and implementing positive training for caregivers. Digital behavioral change information campaigns that aim to build positive relationships between parent and child, prevent misbehavior, manage misbehavior, and improve emotional self-regulation can help to increase the quality of parenting. Moreover, this learning can help reduce harmful parenting practices during the early ages of the children, diminishing exposure to violence at home (Francis and Baker-Henningham, 2021). Altogether, a more positive and healthy

[17]For further detail, Table A15 shows detailed cost information (fixed and variable) associated with our intervention as well as the face-to-face program.

rearing environment has the potential to increase the long-term well-being of children and their families.

References

Aguero, J. and Frisancho, V. (2021). Measuring violence against women with experimental methods. *Economic Development and Cultural Change*.

Altafim, E. R. P. and Linhares, M. B. M. (2016). Universal violence and child maltreatment prevention programs for parents: A systematic review. *Psychosocial Intervention*, 25(1):27–38.

Amaral, S., Dinarte Diaz, L., Dominguez, P., and Perez-Vincent, S. M. (2021). Helping families help themselves? Heterogeneous effects of a digital parenting program.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Asadullah, M. N., De Cao, E., Khatoon, F. Z., and Siddique, Z. (2021). Measuring gender attitudes using list experiments. *Journal of Population Economics*, 34(2):367–400.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148 – 1178.

Athey, S. and Wager, S. (2019). Estimating treatment effects with causal forests: An application.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: Results from a randomized controlled trial in Colombia. *American Economic Review*, 110(1):48–85.

Attanasio, O. P., Fernández, C., Fitzsimons, E. O., Grantham-McGregor, S. M., Meghir, C., and Rubio-Codina, M. (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in colombia: cluster randomized controlled trial. *Bmj*, 349.

Baker-Henningham, H. and Francis, T. (2018). Parents' use of harsh punishment and young children's behaviour and achievement: a longitudinal study of jamaican children with conduct problems. *Global Mental Health*, 5.

Balaji, A. B., Claussen, A. H., Smith, D. C., Visser, S. N., Morales, M. J., and Perou, R. (2007). Social support networks and maternal mental health and well-being. *Journal of women's health*, 16(10):1386–1396.

Baranov, V., Bhalotra, S., Biroli, P., and Maselko, J. (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review*, 110(3):824–59.

Barrera, O., Macours, K., Premand, P., and Vakis, R. (2020). Texting parents about early child development. *World Bank Policy Research Working Paper*, (9492).

Bell, K. M. and Naugle, A. E. (2007). Effects of social desirability on students' self-reporting of partner abuse perpetration and victimization. *Violence and victims*, 22(2):243–256.

Belsky, J. (1984). The determinants of parenting: A process model. *Child development*, pages 83–96.

Belsky, J. and Jaffee, S. R. (2006). The multiple determinants of parenting.

Bendini, M. and Dinarte, L. (2020). Does maternal depression undermine childhood cognitive development? Evidence from the Young Lives Survey in Peru. *International Journal of Environmental Research and Public Health*, 17(19):7248.

Berry, J. O. and Jones, W. H. (1995). The parental stress scale: Initial psychometric evidence. *Journal of social and personal relationships*, 12(3):463–472.

Blattman, C., Green, D. P., Ortega, D., and Tobón, S. (2021). Place-based interventions at scale: The direct and spillover effects of policing and city services on crime. *Journal of the European Economic Association*, 19(4):2022–2051.

Boden, J. M., Horwood, L. J., and Fergusson, D. M. (2007). Exposure to childhood sexual and physical abuse and subsequent educational achievement outcomes. *Child abuse & neglect*, 31(10):1101–1114.

Bos, J., Shonchoy, A., Ravindran, S., and Khan, A. (2022). Early childhood human capital formation at scale. *Available at SSRN 3906697*.

Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.

Carneiro, P., Galasso, E., García, I. L., Bedregal, P., and Cordero, M. (2023). Parental beliefs, investments, and child development: Evidence from a large-scale experiment. *Journal of Political Economy*, forthcoming.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1):211–28.

Chandra, A., Mani, S., Dolphin, H., Dyson, M., and Marah, Y. (2021). Experimental evidence from an integrated early childhood parenting program in sierra leone.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2023). Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India. *NBER Working Paper*, (w24678).

Chorpita, B. F. and Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of consulting and clinical psychology*, 77(3):566.

Cluver, L. D., Meinck, F., Steinert, J. I., Shenderovich, Y., Doubt, J., Romero, R. H., Lombard, C. J., Redfern, A., Ward, C. L., Tsoanyane, S., Nzima, D., Sibanda, N., Wittesaele, C., Stone, S. D., Boyes, M. E., Catanho, R., Lachman, J. M., Salah, N., Nocuza, M., and Gardner, F. (2018). Parenting for Lifelong Health: a pragmatic cluster randomised controlled trial of a non-commercialised parenting programme for adolescents and their families in South Africa. *BMJ Global Health*, 3(1).

Cochran, M. (2019). *Chapter 6: Parenting and Social Networks*. Handbook of Parenting: Volume 4 Social Conditions and Applied Parenting.

Cortes, K. E., Fricke, H. D., Loeb, S., Song, D. S., and York, B. N. (2019). When behavioral barriers are too high or low–how timing matters for parenting interventions. Technical report, National Bureau of Economic Research.

Crowne, D. P. and Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4):349.

Currie, J. and Spatz Widom, C. (2010). Long-term consequences of child abuse and neglect on adult economic well-being. *Child maltreatment*, 15(2):111–120.

Currie, J. and Tekin, E. (2012). Understanding the cycle childhood maltreatment and future crime. *Journal of Human Resources*, 47(2):509–549.

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–74.

de Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302.

de Simone, F., Mejia, C., Martinez-Carrasco, J., Perez-Vincent, S. M., and Villalba, H. (2022). Mitigating Coercive Parenting Through Home Visitations: Impacts of a Parenting Program Targeted at Vulnerable Communities in Jamaica. *IDB Working Paper*.

Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American Economic Review*, 112(3):899–927.

Donker, T., van Straten, A., Marks, I., and Cuijpers, P. (2011). Quick and easy self-rating of generalized anxiety disorder: validity of the dutch web-based gad-7, gad-2 and gad-si. *Psychiatry research*, 188(1):58–64.

Doss, C., Fahle, E. M., Loeb, S., and York, B. N. (2019). More than just a nudge supporting kindergarten parents with differentiated and personalized text messages. *Journal of Human Resources*, 54(3):567–603.

Doyle Jr, J. J. and Aizer, A. (2018). Economics of child protection: Maltreatment, foster care, and intimate partner violence. *Annual review of economics*, 10:87–108.

Fang, Y., Boelens, M., Windhorst, D. A., Raat, H., and van Grieken, A. (2021). Factors associated with parenting self-efficacy: A systematic review. *Journal of Advanced Nursing*, 77(6):2641–2661.

Fernández-González, L., O'Leary, K. D., and Muñoz-Rivas, M. J. (2013). We are not joking: Need for controls in reports of dating violence. *Journal of Interpersonal violence*, 28(3):602–620.

Francis, T. and Baker-Henningham, H. (2020). Design and implementation of the irie homes toolbox: a violence prevention, early childhood, parenting program. *Frontiers in public health*, 8:702.

Francis, T. and Baker-Henningham, H. (2021). The Irie Homes Toolbox: A cluster randomized controlled trial of an early childhood parenting program to prevent violence against children in Jamaica. *Children and Youth Services Review*, 126:106060.

Fry, D., Fang, X., Elliott, S., Casey, T., Zheng, X., Li, J., Florian, L., and McCluskey, G. (2018). The relationships between violence in childhood and educational outcomes: A global systematic review and meta-analysis. *Child abuse & neglect*, 75:6–28.

Garland, A. F., Hawley, K. M., Brookman-Frazee, L., and Hurlburt, M. S. (2008). Identifying common elements of evidence-based psychosocial treatments for children's disruptive behavior problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(5):505–514.

Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40.

Hamadani, J. D., Mehrin, S. F., Tofail, F., Hasan, M. I., Huda, S. N., Baker-Henningham, H., Ridout, D., and Grantham-McGregor, S. (2019). Integrating an early childhood development programme into bangladeshi primary health-care services: an open-label, cluster-randomised controlled trial. *The Lancet Global Health*, 7(3):e366–e375.

Hamby, S., Finkelhor, D., Turner, H., and Ormrod, R. (2011). *Children's Exposure to Intimate Partner Violence and Other Family Violence*. Washington, DC: U.S. Department of Justice.

Heath, R., Hidrobo, M., and Roy, S. (2020). Cash transfers, polygamy, and intimate partner violence: Experimental evidence from mali. *Journal of Development Economics*, 143(102410).

Heckman, J. J., Liu, B., Lu, M., and Zhou, J. (2020). The impacts of a prototypical home visiting program on child skills. *NBER Working Paper*, (w27356).

Hoover-Dempsey, K. V., Walker, J. M., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., and Closson, K. (2005). Why do parents become involved? research findings and implications. *The elementary school journal*, 106(2):105–130.

Hurwitz, L. B., Lauricella, A. R., Hanson, A., Raden, A., and Wartella, E. (2015). Supporting head start parents: impact of a text message intervention on parent–child activity engagement. *Early Child Development and Care*, 185(9):1373–1389.

Izzo, C., Weiss, L., Shanahan, T., and Rodriguez-Brown, F. (2000). Parental self-efficacy and social support as predictors of parenting practices and children's socioemotional adjustment in mexican immigrant families. *Journal of Prevention & Intervention in the Community*, 20(1-2):197–213.

Jones, T. L. and Prinz, R. J. (2005). Potential roles of parental self-efficacy in parent and child adjustment: A review. *Clinical psychology review*, 25(3):341–363.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Knerr, W., Gardner, F., and Cluver, L. (2013). Improving Positive Parenting Skills and Reducing Harsh and Abusive Parenting in Low- and Middle-Income Countries: A Systematic Review . *Prevention Science*, 14:352–363.

Kroenke, K., Spitzer, R. L., and Williams, J. B. (2003). The patient health questionnaire-2: validity of a two-item depression screener. *Medical care*, pages 1284–1292.

Lachman, J., Wamoyi, J., Spreckelsen, T., Wight, D., Maganga, J., and Gardner, F. (2020). Combining parenting and economic strengthening programmes to reduce violence against children: a cluster randomised controlled trial with predominantly male caregivers in rural Tanzania. *BMJ Global Health*, 5(7).

Landry, S. H., Smith, K. E., Swank, P. R., Zucker, T., Crawford, A. D., and Solari, E. F. (2012). The effects of a responsive parenting intervention on parent–child interactions during shared book reading. *Developmental psychology*, 48(4):969.

Lansford, J. E., Dodge, K. A., Pettit, G. S., Bates, J. E., Crozier, J., and Kaplow, J. (2002). A 12-year prospective study of the long-term effects of early child physical maltreatment on psychological, behavioral, and academic problems in adolescence. *Archives of pediatrics & adolescent medicine*, 156(8):824–830.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102.

Levere, M., Acharya, G., and Bharadwaj, P. (2016). The role of information and cash transfers on early childhood development: evidence from nepal. Technical report, National Bureau of Economic Research.

Macours, K., Schady, N., and Vakis, R. (2012). Cash transfers, behavioral changes, and cognitive development in early childhood: evidence from a randomized experiment. *American Economic Journal: Applied Economics*, 4(2):247–73.

Mehrin, S. F., Hasan, M. I., Tofail, F., Shiraji, S., Ridout, D., Grantham-McGregor, S., Hamadani, J. D., and Baker-Henningham, H. (2022). Integrating a group-based, early childhood parenting intervention into primary health care services in rural bangladesh: A cluster-randomised controlled trial. *Frontiers in Pediatrics*.

Mersky, J. P. and Topitzes, J. (2010). Comparing early adult outcomes of maltreated and non-maltreated children: A prospective longitudinal investigation. *Children and Youth Services Review*, 32(8):1086–1096.

Michie, S., Van Stralen, M. M., and West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation science*, 6(1):1–12.

Michl-Petzing, L. C., Handley, E. D., Sturge-Apple, M., Cicchetti, D., and Toth, S. L. (2019). Re-examining the "cycle of abuse": Parenting determinants among previously maltreated, low-income mothers. *Journal of family psychology*, 33(6):742.

Operator Watch (2021). Jamaica has good 4G coverage but no certainty on 5G. Available at https://www.operatorwatch.com/2021/04/jamaica-has-good-4g-coverage-but-no.html.

Owen, A. and Anderson, B. (2017). Informal community support for parents of pre-school children: A comparative study investigating the subjective experience of parents attending community-based toddler groups in different socio-economic situations. *Journal of Early Childhood Research*, 15(2):212–224.

Paxson, C. and Schady, N. (2010). Does money matter? the effects of cash transfers on child development in rural ecuador. *Economic development and cultural change*, 59(1):187–229.

Persson, P. and Rossin-Slater, M. (2018). Family ruptures, stress, and the mental health of the next generation. *American Economic Review*, 108(4-5):1214–52.

Smith, J. A., Chang, S. M., Brentani, A., Fink, G., Lopez-Boo, F., Torino, B. M., Codina, M. R., and Walker, S. P. (2023). A Remote Parenting Program and Parent and Staff Perspectives: A Randomized Trial. *Pediatrics*, 151(Supplement 2). e2023060221F.

Sviatschi, M. M. (2022). Making a narco: Childhood exposure to illegal labor markets and criminal life paths. *Econometrica*, 90(4):1835–1878.

Szekely, M. (2023). Results from focus group discussions with participants of the program Virtual Irie Homes in Jamaica. Technical report.

Taraban, L. and Shaw, D. S. (2018). Parenting in context: Revisiting belsky's classic process of parenting model in early childhood. *Developmental Review*, 48:55–81.

UNICEF (2011). 2011 Multiple Indicator Cluster Survey. Technical report, UNICEF.

UNICEF (2017). A familiar face: Violence in the lives of children and adolescents. *New York, NY: UNICEF*.

UNICEF (2022). A statistical profile of violence against children in latin america and the caribbean. Technical report, United Nations Children's Fund (UNICEF).

Widen, S. C., Orozco, M., Horng, E. L., and Loeb, S. (2020). Reaching unconnected caregivers: Using a text-message education program to better understand how to support informal caregivers roles in child development. *Journal of Early Childhood Research*, 18(1):29–43.

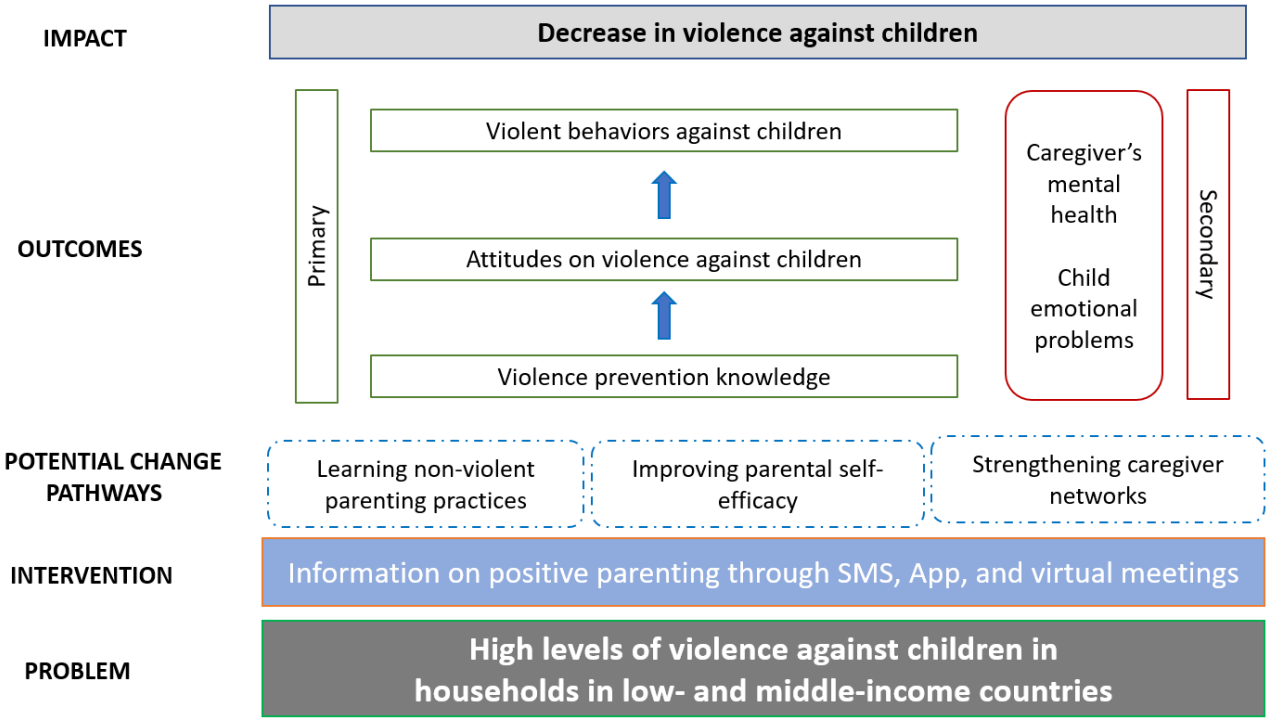
Wittkowski, A., Garrett, C., Calam, R., and Weisberg, D. (2017). Self-report measures of parental self-efficacy: A systematic review of the current literature. *Journal of child and family studies*, 26:2960–2978.

York, B. N., Loeb, S., and Doss, C. (2019). One step at a time the effects of an early literacy text-messaging program for parents of preschoolers. *Journal of Human Resources*, 54(3):537–566.

Tables and Figures

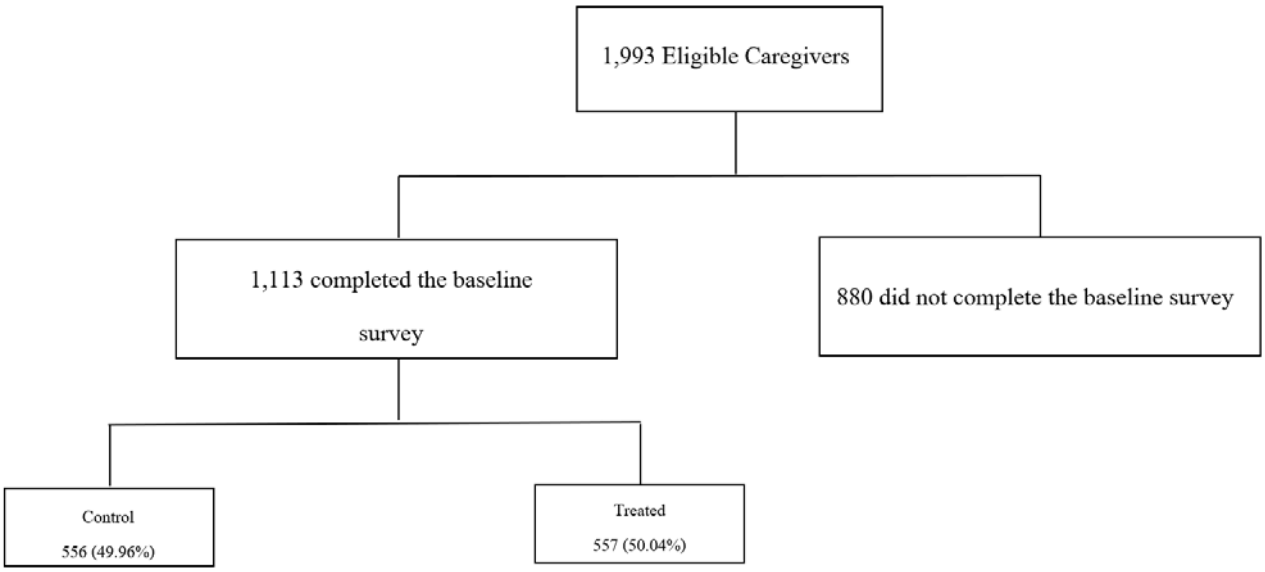
Figures

Figure 1: Conceptual Framework



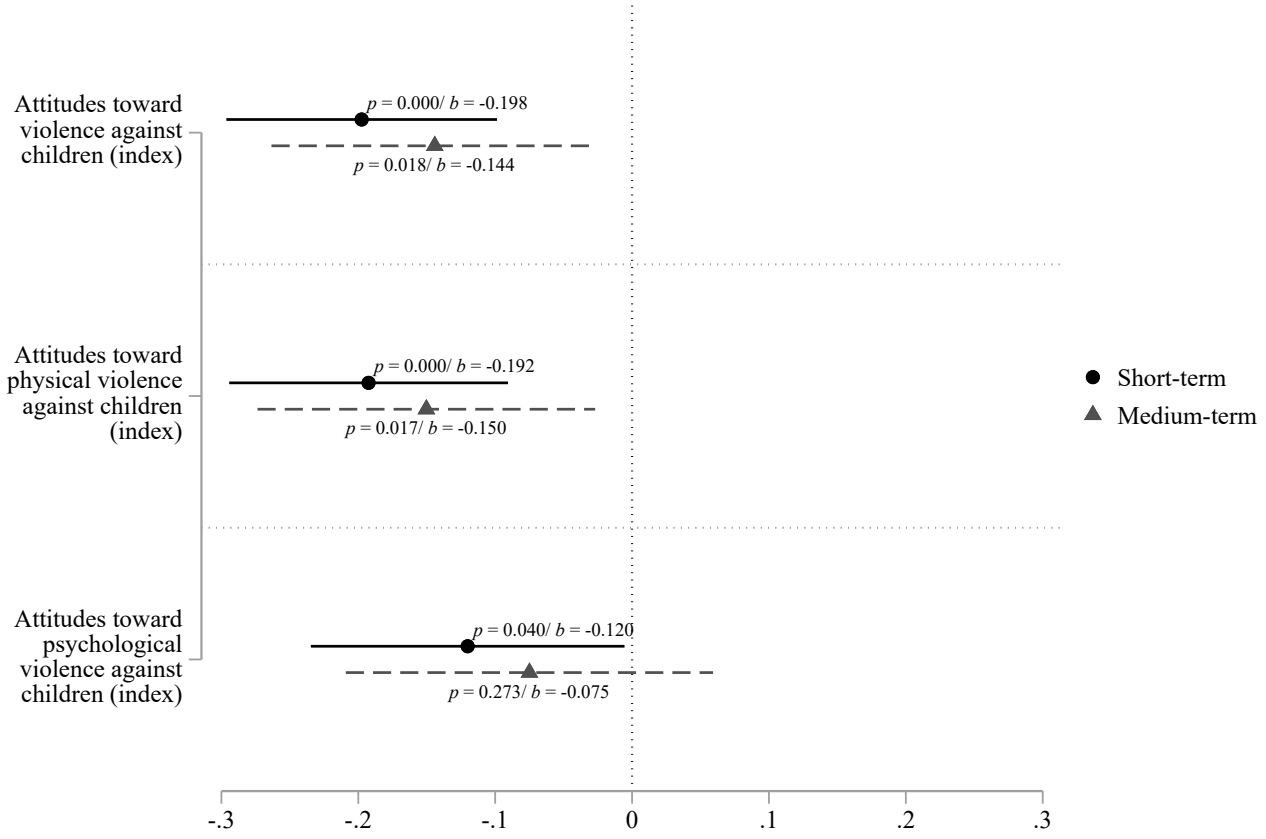
Notes: This diagram shows the conceptual framework for our analysis. For more details about this framework, see the text.

Figure 2: Experimental Design



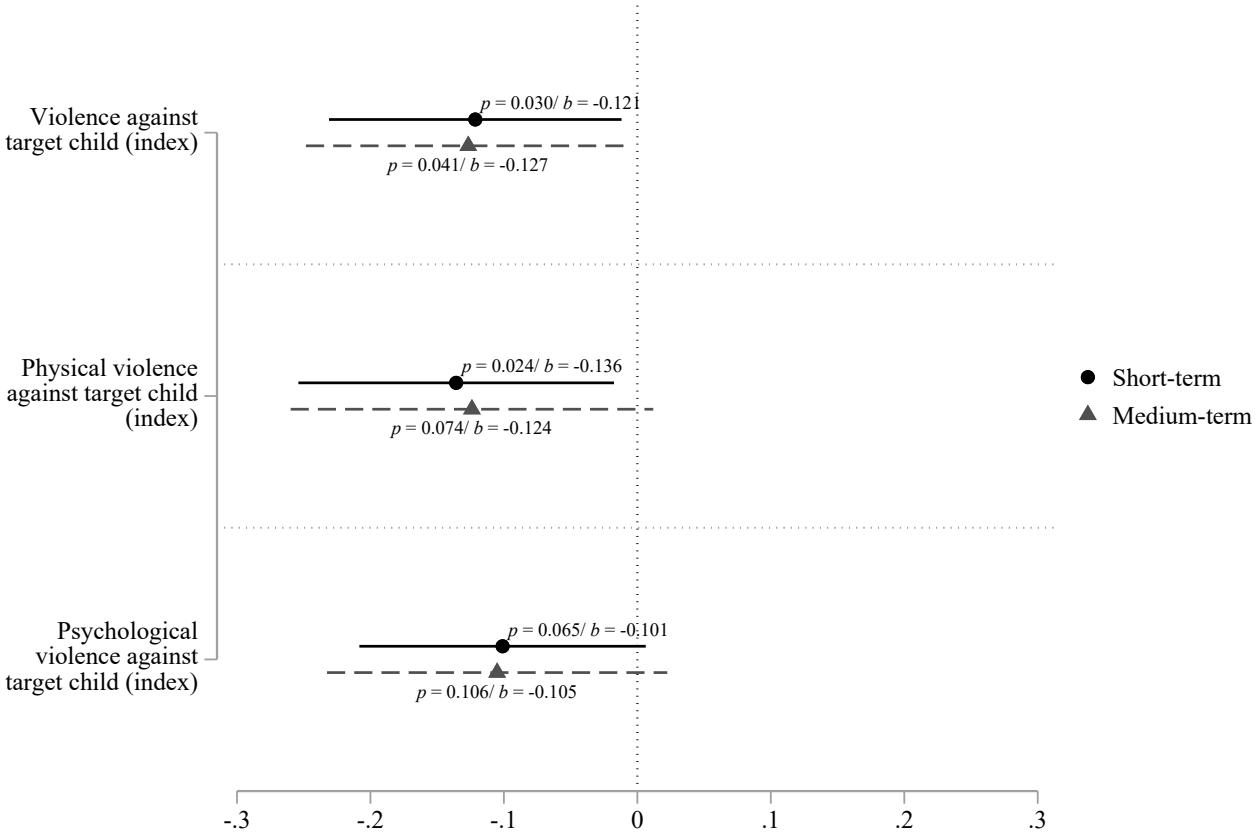
Notes: This figure summarizes the experimental design of the study. 880 enrolled individuals did not complete the baseline survey for several reasons, including they did not provide a correct phone number; we were unable to reach them after the maximum number of attempts determined in the ethics protocol; they changed their mind and decided not to participate in the study, among others.

Figure 3: ITT Impacts on Caregiver Attitudes Toward Violence Against Children



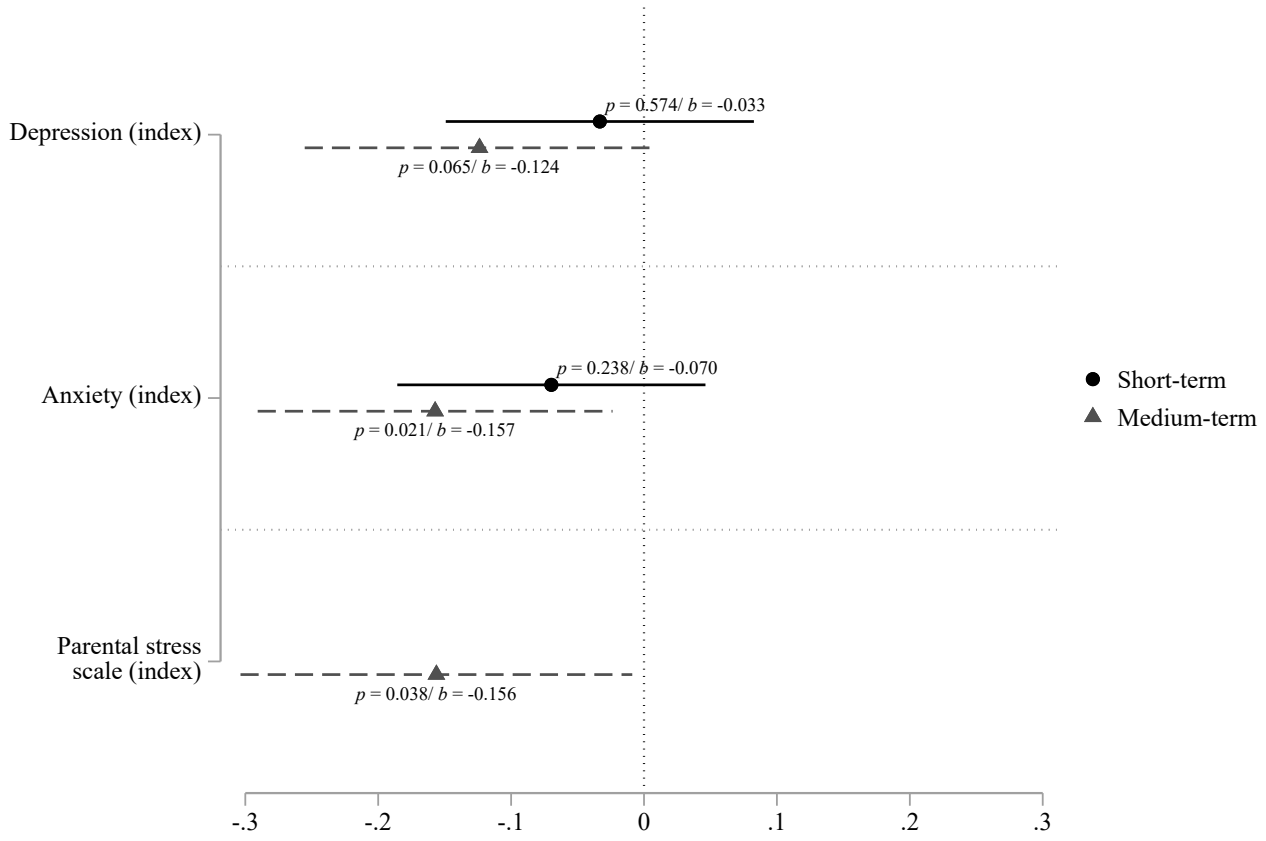
Notes: This figure presents estimates of β_1 (the ITT estimate, b) from Equation 1 (and the respective p-value) on caregivers' attitudes toward violence against children. The black circles and corresponding solid lines represent the point estimates and 95% confidence intervals from the first follow-up survey (short-term). The gray triangles and corresponding dashed lines represent the point estimates and 95% confidence intervals from the second follow-up survey (medium-term). Each outcome consists of a standardized index estimated following Anderson (2008) and standardized relative to the control group. For a detailed description of the indices, see Section 3.3. All specifications include controls for strata fixed effects. Standard deviation units are used for the x-axis.

Figure 4: ITT Impacts on Caregiver Behaviors Related to Violence Against the Target Child



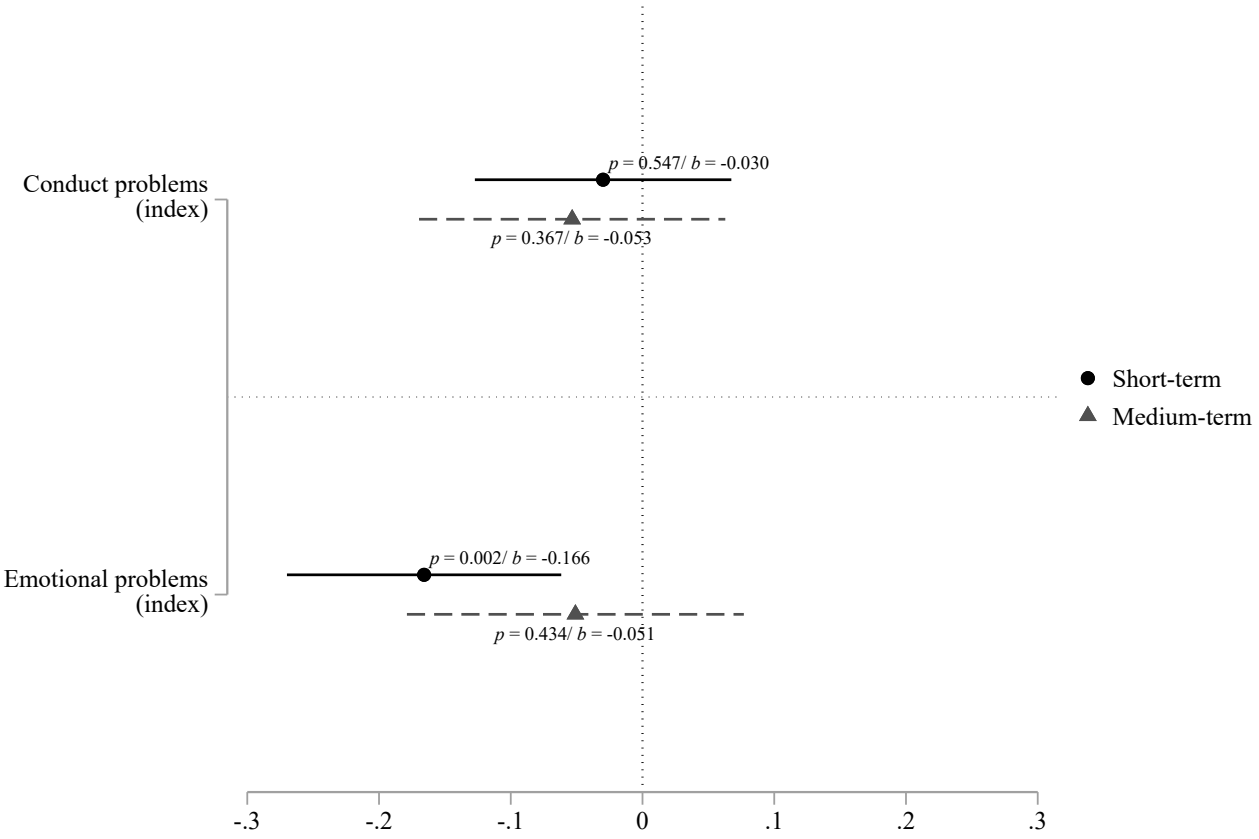
Notes: This figure presents estimates of β_1 (the ITT estimate, b) from Equation 1 (and the respective p-value) on caregivers' behaviors relating to violence against the target child. The black circles and corresponding solid lines represent the point estimates and 95% confidence intervals from the first follow-up survey (short-term). The gray triangles and corresponding dashed lines represent the point estimates and 95% confidence intervals from the second follow-up survey (medium-term). Each outcome consists of a standardized index estimated following Anderson (2008) and standardized relative to the control group. For a detailed description of the indices, see Section 3.3. All specifications include controls for strata fixed effects. Standard deviation units are used for the x-axis.

Figure 5: ITT Impacts on Caregivers' Mental Health



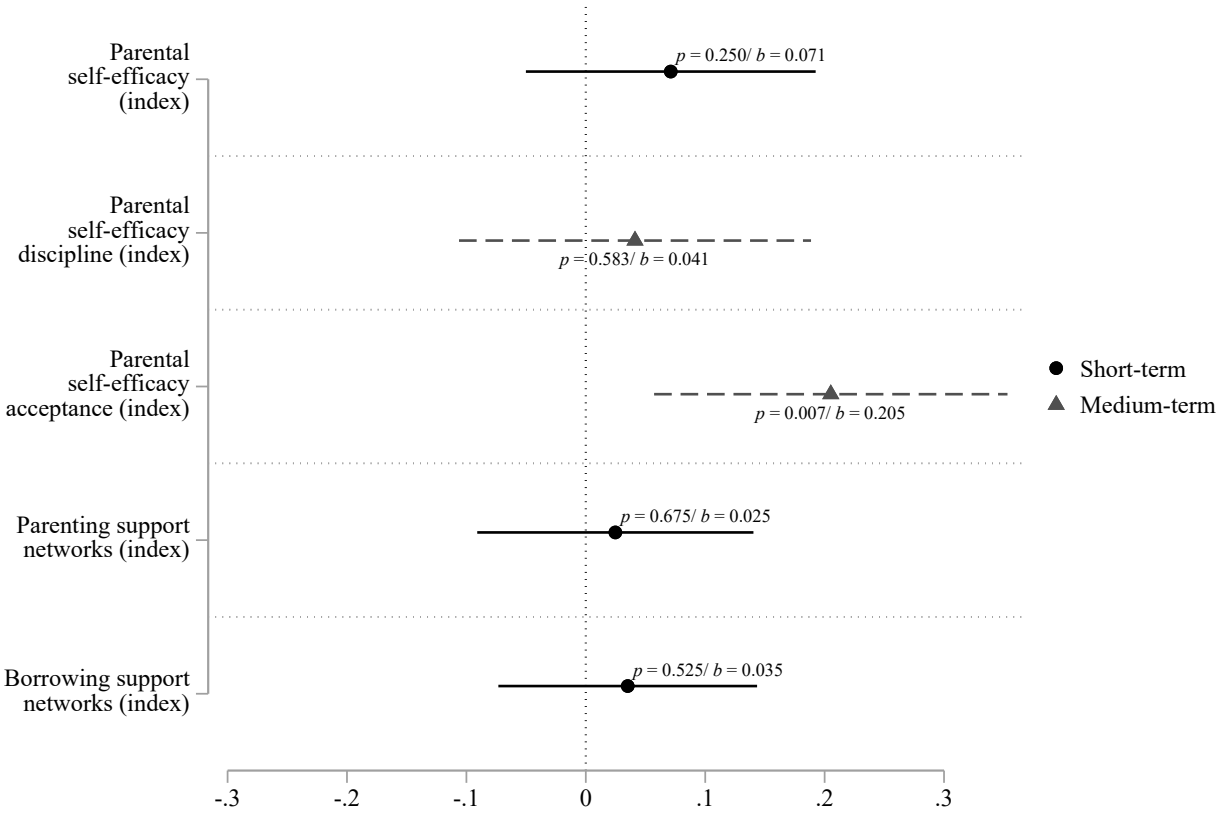
Notes: This figure presents estimates of β_1 (the ITT estimate, b) from Equation 1 (and the respective p-value) on caregivers' depression, anxiety, and parental stress. The black circles and corresponding solid lines represent the point estimates and 95% confidence intervals from the first follow-up survey (short-term). The gray triangles and corresponding dashed lines represent the point estimates and 95% confidence intervals from the second follow-up survey (medium-term). Caregivers' depression and anxiety were measured using PHQ-2 and GAD-2 at both follow-ups, respectively. Parental stress was measured using PSS-18 only at the second follow-up (medium-term). Each outcome consists of a standardized index estimated following Anderson (2008) and standardized relative to the control group. For a detailed description of the indices, see Section 3.3. All specifications include controls for strata fixed effects. Standard deviation units are used for the x-axis.

Figure 6: ITT Impacts on Child Conduct and Emotional Problems



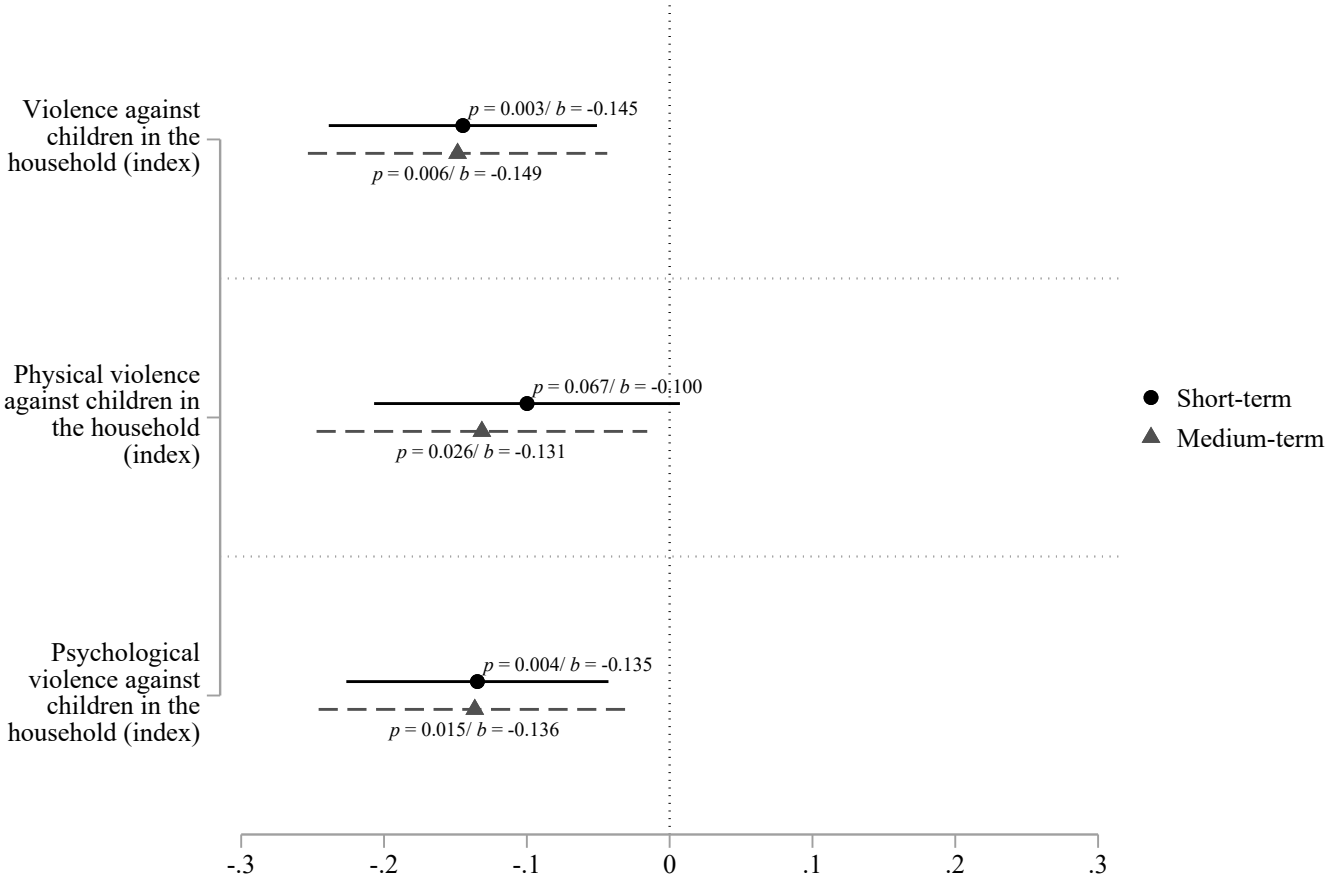
Notes: This figure presents estimates of β_1 (the ITT estimate, b) from Equation 1 (and the respective p-value) on the target child's conduct and emotional problems. The black circles and corresponding solid lines represent the point estimates and 95% confidence intervals from the first follow-up survey (short-term). The gray triangles and corresponding dashed lines represent the point estimates and 95% confidence intervals from the second follow-up survey (medium-term). Conduct and emotional problems were measured using caregivers' responses to the Strengths and Difficulties Questionnaire (SDQ) instrument. Each outcome consists of a standardized index estimated following Anderson (2008) and standardized relative to the control group. For a detailed description of the indices, see Section 3.3. All specifications include controls for strata fixed effects. Standard deviation units are used for the x-axis.

Figure 7: Potential Mechanisms



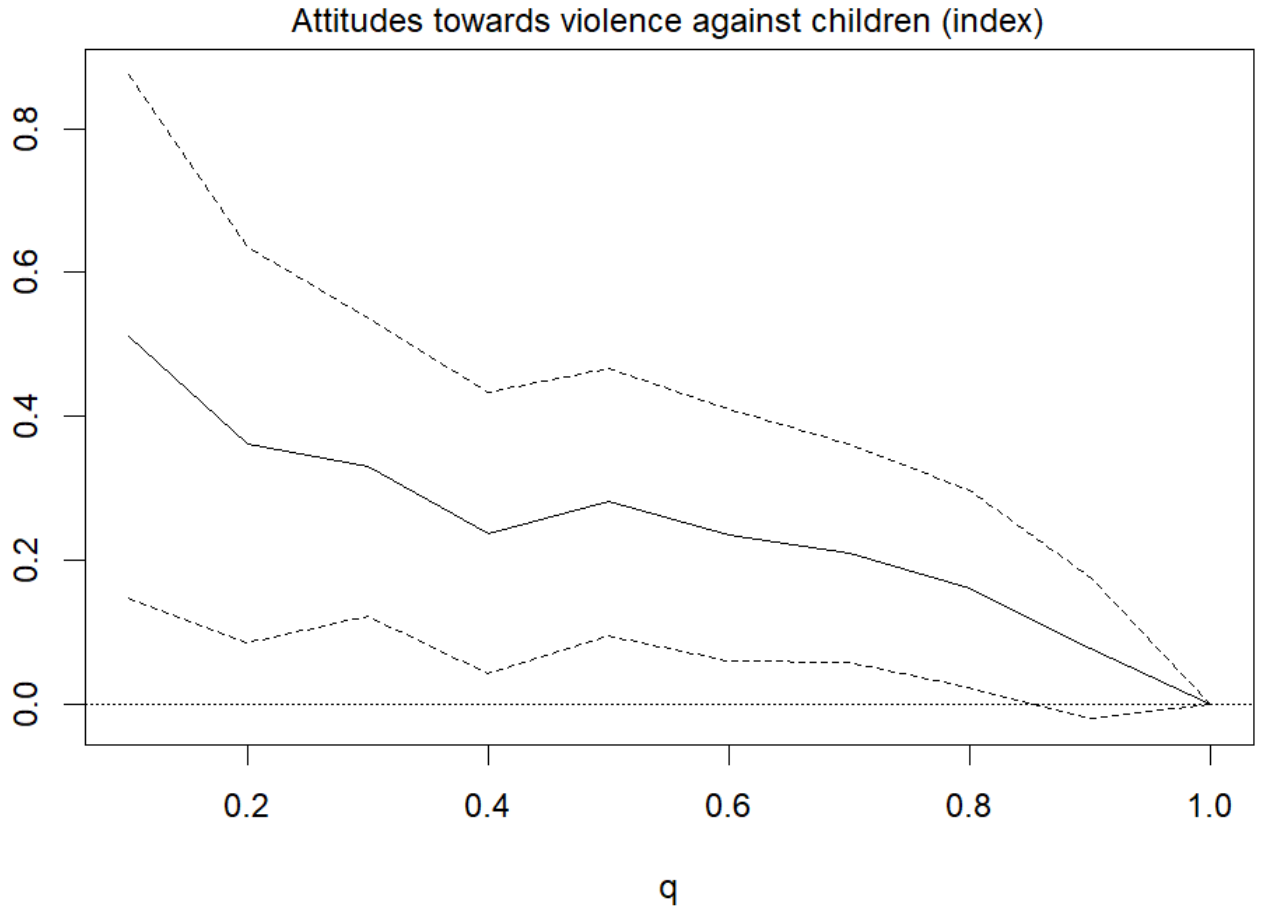
Notes: This figure presents estimates of β_1 (the ITT estimate, b) from Equation 1 (and the respective p-value) on five variables that explore potential mechanisms through which the intervention improved the caregivers' self-efficacy and support networks. The black circles and corresponding solid lines represent the point estimates and 95% confidence intervals from the first follow-up survey (short-term). The gray triangles and corresponding dashed lines represent the point estimates and 95% confidence intervals from the second follow-up survey (medium-term). To measure self-efficacy, we used the Brief Parental Self-Efficacy Scale (BPSES) at the first follow-up but adapted to the Tool to Measure Parenting Self-Efficacy (TOPSE) at the second follow-up for more detailed questions relating to discipline and self-acceptance. For networks, we examined parenting support networks and borrowing support networks. Each outcome consists of a standardized index estimated following Anderson (2008) and standardized relative to the control group. For a detailed description of the indices, see Section 3.3. All specifications include controls for strata fixed effects. Standard deviation units are used for the x-axis.

Figure 8: ITT Impacts on Caregiver Behaviors Related to Violence Against Other Children in Household



Notes: This figure presents estimates of β_1 (the ITT estimate, b) from Equation 1 (and the respective p-value) on caregivers' behaviors relating to violence against the target child and the eldest child (if any, aged 7-12) in the household. The black circles and corresponding solid lines represent the point estimates and 95% confidence intervals from the first follow-up survey (short-term). The gray triangles and corresponding dashed lines represent the point estimates and 95% confidence intervals from the second follow-up survey (medium-term). Each outcome consists of a standardized index estimated following Anderson (2008) and standardized relative to the control group. For a detailed description of the indices, see Section 3.3. All specifications include controls for strata fixed effects. Standard deviation units are used for the x-axis.

Figure 9: Heterogeneity in Treatment Impacts: Machine Learning Approach



Notes: This figure presents the Targeting Operator Characteristic for the outcome attitudes toward violence against children (index). Data from the short- and medium-term follow-ups have been pooled for statistical power. The x-axis denotes the top q-th fraction of individuals with the largest prioritization score, i.e. caregivers believed to have the largest benefit from the treatment as given by the Conditional Average Treatment Effect (CATE). The TOC compares the ATE in smaller groups defined by the prioritization rule to the overall ATE from treating everyone in the treatment group. The CATEs are estimated non-parametrically using the $causal\_forest$ and $rank\_average\_treatment\_effect$ functions in R. 95% confidence intervals are shown. Standard deviation units are used for the y-axis.

Tables

Table 1: Summary Statistics by Group and Balance Tests

	Control		Treatment		*p*-value
	(1)	(2)	(3)	(4)	(5)
Variable	Mean	SD	Mean	SD	(1) - (3)
Panel A. Caregiver's characteristics					
Age (years)	33.405	7.727	33.070	7.247	0.443
Female (%)	0.856	0.351	0.853	0.355	0.634
Education level completed (years)	14.220	2.761	14.445	2.733	0.166
Married (%)	0.384	0.487	0.356	0.479	0.450
Employed (%)	0.784	0.412	0.792	0.407	0.700
Income in the past month (USD)	910.306	1,166.991	855.201	1,062.471	0.489
Household size (N)	4.559	1.873	4.598	2.025	0.734
Children 17 years or younger (N)	1.950	1.021	1.873	1.044	0.213
Panel B. Target child's characteristics					
Age (years)	4.171	1.429	4.070	1.425	0.236
Female (%)	0.480	0.500	0.496	0.500	0.623
Panel C. Primary Outcomes					
Attitudes toward violence against children (% in agreement)					
Shouting and yelling makes the child more obedient	0.096	0.295	0.115	0.320	0.311
Shouting, yelling, and threatening to slap will not harm the child	0.368	0.483	0.362	0.481	0.830
To raise a child properly, the child needs to be physically punished	0.115	0.319	0.131	0.338	0.391
A good parent slaps their child when they misbehave	0.203	0.403	0.204	0.404	0.941
When a child is beaten, he/she will stop doing the unwanted behavior	0.097	0.296	0.098	0.297	0.948
Violence against target child (Number of days in the past 7 days)					
Shouted, yelled, or screamed at him/her?	1.541	1.394	1.632	1.352	0.270
Said you would send him/her away?	0.228	0.719	0.142	0.506	0.023
Threatened to hit him/her but not actually done it?	1.479	1.498	1.540	1.485	0.511
Hit him/her on the bottom, hand, arm, or leg with your bare hand?	0.519	0.790	0.525	0.728	0.926
Hit him/her on the bottom, hand, arm, or leg with a hard object?	0.081	0.356	0.097	0.393	0.482
Panel D. Secondary Outcomes					
Conduct problems (%)	41.655	27.063	41.939	26.966	0.849
Emotional problems (%)	24.676	23.305	25.458	22.925	0.566
Depression (%)	19.964	40.009	18.133	38.564	0.440
Generalized Anxiety Disorder (%)	13.309	33.998	17.056	37.646	0.084
Panel E. Mechanisms					
Borrowing money support networks (N)	2.302	2.555	2.336	2.490	0.825
Parenting issues support networks (N)	2.558	2.691	2.743	2.680	0.247
Belong to parent support group (%)	0.155	0.362	0.162	0.368	0.761
F-test of joint significance (p-value)					0.968

Notes: This table shows average characteristics at baseline for the study participants assigned to the treatment and control groups. Columns (1) and (2) present the mean and standard errors of the variables for the control group, while columns (3) and (4) present the mean and standard errors of the variables for the treatment group, respectively. Column (5) shows the *p*-value associated with the hypothesis of the mean values across both groups being the same. We imputed the mean to have consistent sample sizes in the following variables: **Panel E:** Borrowing money support networks (N), Parenting issues support networks (N), Belong to parent support group (%).

Table 2: First Stage: ITT Impacts on Learning

	Information module								
	(1) Praising children helps	(2) Important for parents to play with child	(3) Clear instructions help	(4) Understand why child misbehaves	(5) Calm down before disciplining	(6) Withdraw attention from child's whining	(7) Redirect rather than reprimand	(8) Consequences and timeout appropriate	(9) Information Index
Panel A: Short-term									
Treatment	0.189***	0.059	0.224***	0.063*	-0.011	0.624***	0.278***	0.080**	0.525***
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.07)	(0.05)	(0.04)	(0.07)
Observations	978	979	978	974	974	971	973	971	979
Control mean	4.13	4.27	3.98	4.29	4.38	2.73	3.74	4.16	-0.00
Panel B: Medium-term									
Treatment	0.104**	0.039	0.248***	-0.058	-0.026	0.480***	0.165***	0.071	0.394***
	(0.05)	(0.04)	(0.06)	(0.04)	(0.04)	(0.08)	(0.06)	(0.04)	(0.08)
Observations	698	699	699	699	698	691	691	697	699
Control mean	4.196	4.290	3.989	4.408	4.450	2.743	3.810	4.164	0.000

Notes: This table presents estimates of β_1 (the ITT estimate) from Equation 1. Each column is a separate dependent variable. Panel A displays the short-term effects and Panel B the medium-term effects. Columns 1 to 8 show the impact of treatment over eight statements relating to parenting practices. Each statement is designed to evaluate the understanding of the four key concepts of the intervention. All outcome variables are Likert-scale variables ranging from 1 (Strongly disagree) to 5 (Strongly agree). Column 9 presents the ITT impacts of the treatment on the information index, which aggregates the eight outcome statements as described in Section 4. All specifications include controls for strata fixed effects. The control mean in Panels A and B refer to the mean of the control group from the first and second round of data, respectively. Number of observations vary across variables due to differences in response rate. Heteroskedasticity-robust standard errors are reported in parentheses below the coefficient estimates. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table 3: ITT Impacts on Caregiver Attitudes & Behaviors

	Primary hypotheses					
	(1) Attitudes toward violence against children (index)	(2) Attitudes toward physical violence against children (index)	(3) Attitudes toward pychological violence against children (index)	(4) Violence against target child (index)	(5) Physical violence against target child (index)	(6) Psychological violence against target child (index)
Panel A: Short-term						
Treatment	-0.198***	-0.192***	-0.120**	-0.121**	-0.136**	-0.101*
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)
Observations	977	974	961	943	920	942
Control mean	-0.000	-0.000	0.000	-0.000	0.000	-0.000
Panel B: Medium-term						
Treatment	-0.144**	-0.150**	-0.075	-0.127**	-0.124*	-0.105
	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
Observations	696	694	685	681	676	681
Control mean	-0.000	-0.000	-0.014	0.000	0.017	0.008

Notes: This table presents estimates of β_1 (the ITT estimate) from Equation 1. Each column is a separate dependent variable. Panel A displays the short-term effects and Panel B the medium-term effects. Columns (1)-(3) present treatment impacts on the caregiver attitudes while columns (4)-(6) present treatment impacts on caregiver behaviors. All dependent variables are index variables, constructed as described in Section 4. All specifications include controls for strata fixed effects and the baseline dependent variable as a control. The control mean in Panels A and B refer to the mean of the control group from the first and second round of data, respectively. Number of observations vary across variables due to differences in response rate. Heteroskedasticity-robust standard errors are reported in parentheses below the coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: ITT Impacts on Caregivers' Mental Health & Child Outcomes

	Caregivers' Mental Health			**Child Outcomes**	
	(1) Depression (index)	(2) Anxiety (index)	(3) Parental stress scale (index)	(4) Conduct problems (index)	(5) Emotional problems (index)
Panel A: Short-term					
Treatment	-0.033	-0.070		-0.030	-0.166***
	(0.06)	(0.06)		(0.05)	(0.05)
Observations	982	982		961	961
Control mean	0.000	-0.000		0.000	0.000
Panel B: Medium-term					
Treatment	-0.124*	-0.157**	-0.156**	-0.053	-0.051
	(0.07)	(0.07)	(0.08)	(0.06)	(0.07)
Observations	699	699	699	685	685
Control mean	0.000	-0.000	0.000	0.000	0.000

Notes: This table presents estimates of β_1 (the ITT estimate) from Equation 1. Each column is a separate dependent variable. Panel A displays the short-term effects and Panel B the medium-term effects. Columns (1)-(3) present treatment impacts on the caregivers' mental health while columns (4)-(5) present treatment impacts on child conduct and emotional problems. The parental stress scale was only measured at second follow-up. All dependent variables are index variables, constructed as described in Section 4. All specifications include controls for strata fixed effects and columns (1), (2), (4), and (5) additionally include the baseline dependent variable as a control. The control mean in Panels A and B refer to the mean of the control group from the first and second round of data, respectively. Number of observations vary across variables due to differences in response rate. Heteroskedasticity-robust standard errors are reported in parentheses below the coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: First Stage: Take-up of Intervention

Variable	Mean	Std. Dev.	Min.	Max.	N
Panel A. SMS Delivery (Admin Data)					
Sent SMS (%)	92.28	6.42	69	98	30
Panel B. SMS/WhatsApp Receipt (Survey Data)					
Received any SMS/WhatsApp (%)	91.38	28.09	0	100	499
Read SMS/WhatsApp if received (%)	96.94	17.23	0	100	458
Found the SMS/WhatsApp useful if read (%)	98.20	13.32	0	100	444
Panel C. App usage (Admin Data)					
Number of sessions accessed	1.04	1.85	0	10	557
Total time in sessions (mins)	6.94	15.58	0	75	557
Panel D. Virtual sessions (Admin Data)					
Number of sessions attended	4.55	3.44	0	10	557

Notes: This table shows descriptive statistics for the take-up of relevant outcomes for each of the three components of the intervention. The table uses survey data on the reception of the SMS messages, App usage data from the phone company, and ECC officer reports of attendance at the virtual sessions. The unit of observation is a message in "Sent SMS (%)" and a treated caregiver in all other variables.

Table 6: Social Desirability Bias Analysis

	Primary hypotheses						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	SDB (index)	Attitudes toward violence against children (index)	Attitudes toward physical violence against children (index)	Attitudes toward psychological violence against children (index)	Violence against target child (index)	Physical violence against target child (index)	Psychological violence against target child (index)
Panel A: Controlling for SDB							
Treatment	-0.016	-0.145**	-0.151**	-0.076	-0.132**	-0.120*	-0.115*
	(0.08)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
SDB (index)		-0.047	-0.050	-0.050	-0.019	-0.027	-0.033
		(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Observations	700	696	694	685	677	672	677
Control Mean	0.000	-0.000	-0.000	-0.014	0.001	0.016	0.009
Panel B: Heterogeneity by SDB							
Treatment		-0.243***	-0.158*	-0.204**	-0.128	-0.103	-0.079
		(0.09)	(0.09)	(0.10)	(0.09)	(0.10)	(0.10)
High SDB Score		-0.223**	-0.128	-0.271***	-0.138	-0.119	-0.112
		(0.09)	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)
Treatment × High SDB Score		0.198*	0.015	0.256*	-0.009	-0.036	-0.073
		(0.12)	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)
Treat + Treat × High SDB score		-0.045	-0.143*	0.053	-0.138*	-0.138	-0.152*
		(0.08)	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)
Observations		696	694	685	677	672	677
Control Mean		-0.000	-0.000	-0.014	0.001	0.016	0.009

Notes: Panel A presents estimates of β_1 (the ITT estimate) from Equation 1, additionally controlling for SDB (index). Panel B presents heterogeneity by SDB and reports estimates of β_1, β_2, and β_3 (the ITT estimates) from the following equation: $Y_{i,t} = \beta_0 + \beta_1 T_i + \beta_2 HighSDB_i + \beta_3 T_i * HighSDB_i + \beta_4 Y_{i,t-1} + \gamma_s + \varepsilon_{i,t}$. "High SDB score" is a dummy variable that takes the value 1 if the SDB score was above the median SDB score for the sample, and 0 otherwise. Each column is a separate dependent variable. Medium-term effects are shown, estimated using data from the second follow-up and baseline. Column (1) presents treatment impacts on the Social Desirability Bias (index). Columns (2)-(7) of this table show our results on caregiver attitudes and behaviors controlling for SDB (index). The SDB score was only measured at second follow-up. The term "Treat + Treat x High SDB score" denotes the total effect of the treatment for those in the treatment group with above-median SDB scores. All dependent variables are index variables, constructed as described in Section 4. All specifications include controls for strata fixed effects, while specifications (2) - (7) additionally include the baseline dependent variable as a control. The estimation sample for columns (5)-(7) uses 4 fewer observations relative to the estimation sample for columns (4)-(6) in Table 3 as SDB measures were not collected for the 4 caregivers. The control mean refers to the mean of the control group from the second follow-up round of data. The sample size in each specification varies according to the number of observations available for each outcome and to the SDB index. Heteroskedasticity-robust standard errors are reported in parentheses below the coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Attrition Analysis

	First follow-up		Second follow-up	
	(1) In endline	(2) In endline	(3) In endline	(4) In endline
Treatment	0.047**	-0.090	0.008	-0.273
	(0.02)	(0.18)	(0.03)	(0.24)
Treatment × Age (years)		0.003		0.004
		(0.00)		(0.00)
Treatment × Female (%)		-0.033		0.050
		(0.07)		(0.09)
Treatment × Education level completed (years)		0.009		0.001
		(0.01)		(0.01)
Treatment × Married (%)		-0.024		-0.038
		(0.05)		(0.07)
Treatment × Employed (%)		0.021		0.057
		(0.05)		(0.07)
Treatment × Income in the past month (USD)		-0.000		0.000
		(0.00)		(0.00)
Treatment × Household size (N)		-0.009		-0.016
		(0.02)		(0.02)
Treatment × Children 17 years or younger (N)		0.004		0.039
		(0.03)		(0.04)
Treatment × Violence against target child (index)		0.007		-0.027
		(0.02)		(0.03)
Treatment × Depression (index)		-0.025		-0.023
		(0.02)		(0.03)
Treatment × Anxiety (index)		-0.010		0.009
		(0.02)		(0.03)
Treatment × Conduct problems (index)		-0.016		-0.017
		(0.02)		(0.03)
Treatment × Emotional problems (index)		0.023		0.014
		(0.02)		(0.03)
Treatment × Attitudes toward violence against children (index)		-0.001		0.015
		(0.02)		(0.03)
Constant	0.843***	0.684***	0.633***	0.377*
	(0.02)	(0.15)	(0.02)	(0.20)
Observations	1113	1113	1113	1113
Q test pvalue		0.897		0.916

Notes: This table uses data from the first and second follow-up rounds to show the differences in attrition between treatment and control groups for each follow-up. We present estimates of β_1 (the ITT estimate) from Equation 1. The dependent variable "In Endline" in all columns is a dummy indicating if a caregiver responded to the follow-up surveys. Models 1 and 3 measure the impact of the treatment on the follow-up survey respondent. Models 2 and 4 measure the impact of any demographic characteristics or outcome variables measured at the baseline on the probability of completing the follow-ups (these variables were also included on their own but their output has been suppressed for brevity). All regressions include strata fixed-effects. Heteroskedasticity-robust standard errors are reported in parenthesis below the coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Conditional Average Treatment Effect Regressions

	Information index			Attitude toward violence against children (index)			Violence against target child (index)		
	(1) Above Median CATE	(2) Below Median CATE	(3) Difference	(4) Above Median CATE	(5) Below Median CATE	(6) Difference	(7) Above Median CATE	(8) Below Median CATE	(9) Difference
Treatment	0.822***	0.490***	0.333	0.05	-0.514***	0.564***	-0.092	-0.212	0.12
	(0.169)	(0.170)	(0.240)	(0.087)	(0.163)	(0.185)	(0.135)	(0.154)	(0.205)
Observations	1080	1080	1080	1079	1079	1079	1044	1044	1044
Control Mean	0.002	0.002	0.002	0.042	0.042	0.042	0.007	0.007	0.007

Notes: This table uses data from the first and second follow-up rounds and presents results of average treatment effects (ATEs) for caregivers based on groups defined by caregivers with high and low conditional average treatment effects (CATEs, estimated using grf model in R) for the three primary outcomes. All dependent variables are index variables, constructed as described in Section 4. Data from the short- and medium-term follow-ups have been pooled for statistical power. Above (below) median groups represent caregivers whose CATEs are greater than (less than) the median estimated CATEs. The sample size in each specification varies according to the number of observations available for each outcome. Heteroskedasticity-robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Heterogeneity in Treatment Impacts: Interaction Terms Approach

	Attitudes toward violence against children (index)				
	(1)	(2)	(3)	(4)	(5)
Panel A: Short-term					
Treatment	-0.241*	-0.213**	-0.099*	-0.094	-0.107
	(0.13)	(0.09)	(0.06)	(0.07)	(0.07)
Treatment × Female caregiver	0.051				
	(0.15)				
Treatment × High income		0.027			
		(0.11)			
Treatment × Worse attitudes toward violence against children			-0.185*		
			(0.10)		
Treatment × More conduct problems				-0.203**	
				(0.10)	
Treatment × More emotional problems					-0.175*
					(0.10)
Observations	977	774	977	977	977
Panel B: Medium-term					
Treatment	-0.058	-0.181**	-0.032	-0.213**	-0.198**
	(0.17)	(0.11)	(0.07)	(0.09)	(0.08)
Treatment × Female caregiver	-0.100				
	(0.18)				
Treatment × High income		0.102			
		(0.14)			
Treatment × Worse attitudes toward violence against children			-0.206*		
			(0.12)		
Treatment × More conduct problems				0.136	
				(0.12)	
Treatment × More emotional problems					0.111
					(0.12)
Observations	696	555	696	696	696

Notes: This table presents treatment heterogeneity of our results for caregiver attitudes toward violence against children (VAC). We present estimates of β_1 and β_3 (the ITT estimates) from the following equation: $Y_{i,t} = \beta_0 + \beta_1 T_i + \beta_2 Var_i + \beta_3 T_i * Var_i + \beta_4 Y_{i,t-1} + \gamma_s + \varepsilon_{i,t}$. The variables ($Var_i$) "high income", "worse attitudes toward VAC", "more conduct problems", and "more emotional problems" are indicator variables equal to one for above-median values of the underlying variables. The variables that were interacted with "treatment" were also included in the regressions, but their output has been suppressed for brevity. Panel A displays the short-term effects and Panel B the medium-term effects. The dependent variable is an index variable, constructed as described in Section 4. All specifications include controls for strata fixed effects and the baseline dependent variable. The control mean in Panels A and B refer to the mean of the control group from the first and second round of data, respectively. The sample size in each specification varies according to the number of observations available for each outcome. Heteroskedasticity-robust standard errors are reported in parentheses below the coefficient estimates. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Cost-Effectiveness Comparison

	(1) Cost per person (USD)	(2) Treatment effect	(3) Type of violence	(4) Cost per 0.13 SD effect (USD)
vIHT	62.4	-0.13 SD	Violence against (target) child	62.40
		-0.12 SD	Physical violence against (target) child	67.60
		-0.11 SD	Psychological violence against (target) child	73.75
IHT	123.95	-0.29 SD	Physical violence against children	55.56
Cash transfer intervention	648.72	-0.13 SD	Physical violence against female partner	648.72
		-0.12 SD	Psychological violence against female partner	702.78
		-0.17 SD	Physical violence against children	496.08

Notes: This table presents a cost effectiveness comparison between the vIHT and other two interventions: face-to-face IHT in Jamaica (Francis and Baker-Henningham, 2021) and a cash transfer intervention in Mali (Heath et al., 2020). Costs for the vIHT excludes the cost of setting up the App (USD 54.21). Column (1) presents the cost of the intervention (in USD) per person. Columns (2) and (3) present the estimated treatment effect on different violence-related outcomes, respectively. Column (4) presents the cost (in USD) per an average effect of 0.13 SD.