Six-Month Follow-Up Evaluation of the *Moments That Matter*® Program in Western Kenya: Evidence of Sustained Outcomes from a Cluster-Randomized Controlled Trial

Joshua Jeong[1], Juliet McCann[1], Alexandria Hadd[1], & Michael Ochieng[2]

[1] Rollins School of Public Health, Emory University, Atlanta, GA, USA
[2] B&M Consult, Nairobi, Kenya

August 31, 2025

ACKNOWLEDGEMENTS

We are grateful for the collaboration and support received from various partners in conducting this research. We thank staff at Episcopal Relief & Development, specifically Dawn Murdock, Prakash Karn, Priscilla Amuah, Geraldine Sicola, and Elisa Waidelich for their support for the research study and coordination across partners. We thank staff at the Anglican Development Services (ADS)-Nyanza, including George Nyamor, Tobias Aulo, Charles Ombonya, Priscillah Ong'era, and ADS-Western, including Eva Chebet and Allan Akwanalo, for their support with study enrollment, community entry, and community mobilization associated with data collection activities. We thank Wilson Ochuka at B&M Consult for his study coordination and oversight of data collection, field work, and interpretation of results, as well as all the research assistants involved in collecting these data at baseline and endline. We acknowledge Alina Bhojani and Zane Maguet at Emory University for research assistance. We appreciate the study participants for their time and willingness to share their experiences with us. We also thank the Conrad N. Hilton Foundation for funding this research and especially Erin McCarthy and Katie Januario at the Hilton Foundation for their support. Finally, this report reflects the views of the authors and may not necessarily represent those of Episcopal Relief & Development, ADS-Nyanza, ADS-Western, or the Conrad N. Hilton Foundation.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report presents findings from a six-month follow-up evaluation of the Moments That Matter (MTM) parenting program in Western Kenya. The evaluation builds on a previously conducted cluster-randomized controlled trial (RCT) that assessed the program's endline or post-intervention impacts on early childhood development (ECD) and caregiver outcomes. While the endline assessment revealed modest but statistically significant improvements in child socioemotional development and various nurturing care and caregiver wellbeing outcomes, this follow-up aimed to examine the sustainability of these effects in the months following program completion.

At six months post-intervention, many of the caregiver-level gains observed at endline were sustained. These include enhanced primary caregiver stimulation in the home, reductions in harsh discipline, increased male caregiver involvement, improvements in caregiver mental health, and reduced intimate partner violence. These outcomes reflect multidimensional benefits across the key pillars of the Nurturing Care Framework, suggesting that MTM's holistic design and community-based delivery approach can support sustained change in caregiving environments. Although caregiver financial stress was reduced at endline, there was no difference between the intervention and control group at the six-month follow-up.

In contrast, the program's impacts on ECD outcomes were limited at follow-up. Among children under 36 months, the only ECD domain that was marginally significant was the CREDI socioemotional development subscale (effect size = 0.21 SD, $p = 0.056$). No significant differences were observed in other CREDI developmental domain scores or for the measure of child behavior using the Strengths and Difficulties Questionnaire that was applied with all children. County-level differences seen at endline diminished over time, although Nyamira continued to show stronger results in a few selected outcomes.

These findings highlight MTM's effectiveness in sustaining short-term improvements in caregiver behaviors, but limited impacts on child development. Results from this 6-month follow-up evaluation reinforce endline conclusions underscoring the need to strengthen the design and delivery of MTM particularly relating to the content on early learning and responsive caregiving to achieve more durable impacts on ECD outcomes, which is the ultimate goal of parenting programs. These insights are critical for ongoing learning, adaptation, and scale-up of MTM and similar interventions in low-resource settings.

BACKGROUND

Parenting interventions are globally recognized as effective for supporting caregivers and promoting early childhood development (ECD) (Jeong, Franchett, et al., 2021). While the magnitude and breadth of their effectiveness vary based on factors such as program content, implementation quality, sociodemographic characteristics of populations, and contextual setting, a strong evidence base affirms their positive impacts on both children and their caregivers globally across low-resource settings (Britto et al., 2017).

However, much of this evidence on the effectiveness of parenting programs has been derived from short-term post-intervention evaluations. While these endline evaluations provide critical insights into the immediate effectiveness of parenting programs, fewer studies have followed up with participants after the program concludes to understand whether effects are sustained, fade out, or even strengthen over time (Jeong, Pitchik, et al., 2021). One of the few long-term evaluations of a parenting program is the Jamaica home visiting study that demonstrated sustained and far-reaching impacts of early psychosocial stimulation on child, adolescent, and adult developmental outcomes, education, mental health, and earnings three decades later (Gertler et al., 2014; Walker et al., 2022). Yet outside of this seminal trial, as more programs have begun to follow their cohorts over time, the more recent studies have consistently observed a decline or "fade out" of program effects appearing or discernable within 12–24 months after program completion (Jeong, Pitchik, et al., 2021).

For example, in India, post-intervention effect sizes on ECD outcomes ranged from 0.24 to 0.32 SD but declined to 0.13 to 0.18 SD at a 15-month follow-up (Meghir et al., 2023). In Pakistan, initial effects ranged from 0.5 to 0.7 SD and dropped to 0.10 to 0.30 SD two years after the intervention ended (Yousafzai et al., 2016). In Rwanda, an effect size of 0.38 SD at endline decreased to 0.21 SD at the three-year follow-up (Justino et al., 2023). Overall, these recent studies suggest that effects on ECD tend to diminish rather than emerge in short-term follow-up periods.

While longer-term studies suggest that effects on ECD tend to diminish over time, less is known about how these changes unfold in the months following a program's end. Few studies have conducted follow-up assessments within the first year post-intervention, leaving gaps in our understanding of when, and for which outcomes, changes may be sustained, fade, or newly emerge. A 6-month follow-up is a relatively short-term follow-up round that can help clarify these dynamics. For outcomes that showed no significant change at endline, this assessment can help determine whether certain effects require more time to materialize. For outcomes that demonstrated promising endline improvements, it offers an early test of the sustainability and durability of those effects.

This report builds on the cluster-randomized controlled trial (RCT) of the Moments That Matter (MTM) program in Western Kenya to explore these questions and see the effects of the program

in the short-term six months after the completion of the program. For the endline results which are in a separate report, when children were on average 30 months of age, we found a significant benefit on children's socioemotional development (0.25 SD) and various positive effects on primary caregiver outcomes, including improvements in parenting practices, reductions in harsh discipline and intimate partner violence victimization, and improvements in caregiver mental health. Interestingly, the benefits on ECD and early learning outcomes were significantly larger in Nyamira than those observed in Vihiga.

To understand the sustainability of these effects, we conducted a 6-month follow-up with the same cohort when children were on average 36 months of age. This report presents findings from that follow-up, focusing on the primary outcomes of ECD as well as secondary and exploratory outcomes related to caregiver behaviors and wellbeing. Our aim is to generate actionable evidence to inform ongoing learning, adaptation, and decision-making for MTM's future implementation and to contribute to the broader understanding of the durability of MTM effects over time.

METHODS

This evaluation was designed and led by Emory University. The data collection was overseen and managed by B&M Consult, a local research firm in Kenya. Below, we present the study methodology in two parts: first, the impact evaluation at endline, followed by the endline process evaluation. The overall study protocol received ethics approvals from Emory University (Protocol #: STUDY00007935) and the Jaramogi Oginga Odinga Teaching and Referral Hospital – Institutional Scientific Ethics Review Committee (Protocol #: ISERC/JOOTRH/684/22). All research participants provided written informed consent.

Impact evaluation

Study design and sampling

We designed a cluster randomized controlled trial to evaluate the effectiveness of a parenting intervention on early child development and caregiving outcomes. Through a consultative process between the program team and local stakeholders, Borabu subcounty in Nyamira county and Luanda subcounty in Vihiga county were selected as the specific project locations based on these sites having some of the poorest maternal and child health indicators. Within each of the two selected subcounties, the research team randomly selected 4 sub-locations and randomly assigned them to the intervention versus control group. After selecting the sublocations in the study, we then randomly selected 5-6 villages within each sublocation. Finally, we enrolled 13 primary caregivers and their index child in each village, matching the intended group size of the program.

Inclusion criteria for eligibility into the research study included: primary caregivers must have a child aged 0-18 months; the household resides within the geographic boundary of the given research sublocations that were selected into the study; and primary caregiver provides informed consent for themselves and their child under the age of 18 months to participate. Exclusion criteria were teenage caregivers under the age of 18 years (i.e., not legally considered adults) or households in which children were older than 18 months of age.

This sampling plan was guided by a formal sample size calculation, which determined that 285 primary caregiver-child dyads were needed per study arm (570 total between the intervention and control groups). This sample size calculation accounted for up to 20% household attrition at endline, which may foreseeably occur due to various reasons such as families relocating or even dropping out. The sample size calculation was powered specifically to detect a 0.25 standard deviation difference in the primary ECD outcome between the intervention and control arms. A 0.25 standard deviation difference is considered the minimum meaningful effect size for intervention benefits on ECD outcomes in the field. A recent systematic review indicates that

parenting programs in LMICs achieve an average effect size of 0.40 standard deviations on ECD outcomes.

Data collection process

Quantitative data from caregivers and children were collected at baseline, and the same primary caregiver-child dyads were reassessed at endline and six months post-intervention. All data collection rounds followed a similar structure, including enumerator training, piloting, and field implementation. We partnered with B&M Consult to hire and co-train 12 enumerators and 2 supervisors, who were divided into two teams of six enumerators and one supervisor each per County. Training sessions covered key topics such as effective interviewing skills, research ethics, electronic data collection, and a detailed review of each module of the survey tool covering various measures on parenting and ECD. The training weeks were divided to include in-class trainings and field piloting, with daily debriefs and feedback led by the Emory team. Baseline training took place in Kisumu from February 15-22, 2023, and baseline data collection was conducted between February 23 and March 10, 2023. Endline training was primarily conducted from October 30-November 12, 2024 with data collection from November 13-25, 2024 and February 26 - March 10, 2025. Training for the 6-month follow-up was conducted from May 12 -16, 2025 with data collection from May 19-30, 2025.

For six-month follow-up all quantitative data were collected using Android mobile devices via ODK or KoboToolbox. Survey administration time was approximately 1.5 hours per household with interviews typically conducted in caregiver's home or in a central location in the community. After the week of in-person data collection, the research team contacted participants who were away to conduct the interview via phone, as they would have been missed otherwise. At the end of each day, supervisors reviewed completed interviews, identified cases requiring follow-ups, and updated the interview trackers to monitor progress. Overall baseline and endline data collection were conducted in close coordination between B&M and ADS project staff in Nyamira and Vihiga to develop a detailed day-by-day plan for effectively reaching caregivers.

A total sample of 595 primary caregiver-child dyads across 51 villages were recruited into the trial and completed baseline assessments. For endline data collection in November 2024 and February – March 2025, we were able to revisit and reassess 525 households from the original trial cohort from whom we had baseline data. This represented 88% of the original trial cohort (see Table 1). The main reasons for missing caregiver interviews or child assessments at endline included caregiver dropout and replacement, temporary absence (e.g., caregiver attending funeral or visiting relatives for the holidays), or unreachability despite multiple attempts by phone and in-person. Slightly more caregivers were lost to follow-up in Vihiga compared to Nyamira and in the intervention group compared to control group.

Table 1. Primary Caregivers from the originally enrolled RCT cohort that were reassessed during endline data collection in November 2024 and February/March 2025.

	Nyamira	Vihiga	Total
Intervention	136/154=88%	133/156=85%	269/310=87%
Control	128/143=90%	128/142=90%	256/285=90%
Total	264/297=89%	261/298=88%	525/595=88%

For six-month follow-up data collection in May 2025, we were able to revisit and reassess 486 primary caregivers from the original trial cohort from whom we had both baseline and endline data. This represented 82% of the original trial cohort (Table 2). The main reasons for missing caregiver interviews or child assessments at six-month follow-up included primary caregiver moved away due to separation or divorce, caregiver or child passed away, caregiver was unreachable, caregiver declined. Between baseline and six-month follow-up, slightly more caregivers were lost to follow-up in Vihiga compared to Nyamira and in the intervention group compared to control group.

Table 2. Caregivers from the originally enrolled RCT cohort that were reassessed during both endline data collection in November 2024 and February/March 2025 and six-month follow-up in May 2025.

	Nyamira	Vihiga	Total
Intervention	128/154=83%	123/156=79%	251/310=81%
Control	116/143=81%	119/142=84%	235/285=83%
Total	244/297=82%	242/298=81%	486/595=82%

Of the 486 primary interviews with caregivers that were previously assessed at baseline and endline, 40 interviews (8%) were conducted via phone (Table 3). The rate of phone interviews was relatively balanced in Nyamira and Vihiga, but higher in intervention than control groups

Table 3. Interviews from six-month follow-up that were completed via phone, of the sample that was also assessed at baseline and endline.

	Nyamira	Vihiga	Total
Intervention	11/128 = 9%	16/123 = 13%	27/251 = 11%

Control	8/116 = 7%	5/119 = 4%	13/235 = 6%
Total	19/244 = 8%	21/242 = 9%	40/486=8%

Outcome measures

Primary outcome

Early child development

The primary outcome of this trial was ECD and was specifically measured using the Caregiver Reported Early Development Instruments (CREDI) long form version. The CREDI is a population-level measure for assessing ECD among children aged 0-35 months based on caregivers' reports of their child's milestones and skills. CREDI scores were calculated following the official CREDI scoring procedure. For the impact evaluation, we used internally age-standardized scores and assessed changes in 6 CREDI outcomes: the overall CREDI score (which combines all 4 domains), the four developmental subdomain scores (cognitive, language, motor, and social-emotional development), as well as the CREDI-mental health subscale which specifically assesses child mental health problems.

As the CREDI can only be scored among children aged 0-35 months, by the time CREDI had been administered at endline and six-month follow-up, some children had "aged-out" of the measure. Specifically, at endline 80% of children received CREDI scores with this rate dropping to 54% by 6 months post-intervention (Table 4).

Table 4. Proportion of final sample followed up at all three time points that received CREDI scores, by data collection point.

	Baseline	Endline	6-Month
Intervention	251/251 = 100%	199/251 = 79%	135/251 = 54%
Control	235/235 = 100%	192/235 = 82%	128/235 = 54%
Total	486/486 = 100%	391/486 = 80%	263/486=54%

At six-month follow-up, in addition to collecting the CREDI from age-eligible children, we measured child development using the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a validated behavioral screening tool for children aged 2–17 years. The measure includes 25 items that assess psychological attributes (some positive, other negatives) across five subscales, each with 5 items: emotional symptoms, conduct problems, hyperactivity/inattention, peer

relationship problems, and prosocial behavior. Each item is rated on a 3-point Likert scale (0 = not true, 1 = sometimes true, 2 = always true).

A total difficulties score (range: 0–40) was calculated by summing the first four subscales, with higher scores indicating greater behavioral or emotional difficulties. We calculated three additional summary scores: internalizing difficulties (sum of emotional symptoms and peer problems subscales; range: 0-20), externalizing difficulties (sum of conduct problems and hyperactivity subscales; range: 0-20), and prosocial behavior (range: 0-10), following recommended scoring approaches (Dickey & Blumberg, 2004; Goodman et al., 2010). Higher scores on internalizing and externalizing scales indicate greater behavioral and emotional difficulties, while higher scores on the prosocial scale reflect more positive social behaviors.

Secondary outcomes

Early learning and responsive caregiving (stimulation, play/learning materials, books)

Caregiver stimulation practices were measured in terms of the number of developmentally enriching activities (e.g., singing, storytelling, praising) the primary caregiver engaged in with the child. The measure was adapted from the Family Care Indicators and comprised 11 items. Primary caregivers also reported on the stimulation activities of a male caregiver if present in the child's life (84% of these male caregivers were the biological fathers of the child). Stimulation scores were calculated separately for the primary caregiver and male caregiver.

Caregivers also reported the variety of play and learning materials available to the child. A total index score was created for the number of different types of play and learning materials present in the household (e.g., home-made toys, store-bought toys, household items, objects in the natural environment), with higher scores signaling more materials. Caregivers also reported the number of children's books at home for the child, which was analyzed as a separate outcome.

Child safety and security (discipline, birth registration)

Caregiver disciplinary practices were assessed using the Child Discipline Module from the Multiple Indicator Cluster Survey, capturing the number of harsh discipline (physical and psychological) and positive discipline by the primary caregiver and the child. Separate binary indicators were created for any violent discipline, physical punishment, psychological aggression, and positive discipline indicating whether caregivers engaged in any such practice. For harsh disciplinary practices, lower scores indicate better outcomes, whereas for positive discipline, higher scores indicate better outcomes. Similar to stimulation practices, primary caregivers reported on their own disciplinary approaches as well as those used by the male caregiver, if present in the child's life.

Birth registration was measured based on the caregiver's report of whether the index child's birth had been registered with the civil authorities.

Psychosocial wellbeing (parenting stress, social support)

Parenting stress was measured using the Parenting Stress Index-Short Form (PSI-SF) Parental Distress subscale (12 items). A total score was calculated, with higher scores indicating greater distress. In addition to the total scores, we created a binary indicator for "high stress" based on the standard PSI-SF scoring guidelines that defines high parenting stress as above the 81st percentile cutoff.

Social support was measured using the Multidimensional Scale of Perceived Social Support (12 items). This scale included a total of 12 items and captured perceived social support from 3 types of individuals: partner/special person, family, and friends. Higher scores indicate greater perceived social support.

Economic empowerment

Primary caregivers reported their total earnings in the past month from all income sources (e.g., salary, casual labor, small businesses) and total current savings (e.g. Bank, SACCO, Chama, Mobile saving). Caregivers who reported having accessed credit in the past month also reported the total amount they accessed in credit.

Male caregiver involvement in household chores

In addition to male caregivers' stimulation and disciplinary practices as already noted above, primary caregivers reported the number of household chores (e.g., washing dishes, cleaning the house) performed by the male caregiver in the past two weeks if he was present in the child's life.

Tertiary Outcomes

Nutrition

Child dietary diversity was measured based on the number of WHO-defined food groups consumed (out of 8) in the previous 24 hours: 1) breast milk; 2) grains, roots and tubers; 3) legumes and nuts; 4) dairy products (milk, yogurt, cheese); 5) flesh foods (meat, fish, poultry, liver or other organs); 6) eggs; 7) vitamin A-rich fruits and vegetables; 8) and other fruits and

vegetables). Higher scores indicated a higher number of food groups consumed by the child in the past 24 hours.

Health

Caregivers reported whether their child had experienced diarrhea, cough, or fever in the past 2 weeks. In addition to the indicators for each of these three illnesses, an overall indicator was created for a child's experience of any illness. Caregivers who reported a child illness were asked if they sought any advice or treatment for that illness. Accordingly, indicators for care-seeking behaviors were created for diarrhea, cough, and fever.

An overall illness indicator was created per standard UNICEF indicator for child illness (diarrhea, cough, or fever). For those who reported a child illness, they were asked if they sought any advice, care, or treatment for that illness. Per UNICEF guidelines, we defined appropriate care-seeking as seeking care from a formal health providers, hospitals, primary health care facilities, or from healthcare workers, including community health promoters.

Intimate partner violence

Female primary caregivers reported intimate partner violence (IPV) victimization by male partners in the past 3 months. These items covered three subscales – physical, emotional, and economic violence – and we also created an overall indicator for any type of IPV victimization. We did not ask about sexual violence as it is the least prevalent of the four types, less directly relevant to the MTM curriculum, and to reduce the overall survey length.

Other outcomes (not prespecified)

Finally, we measured a few additional outcomes that were not pre-specified at the outset of the evaluation in the baseline report or the trial registration. The first was whether caregivers reportedly had a kitchen garden on their compound. The second was community connectedness, which the MTM theory of change posits as a key pathway for improving caregiver well-being. This measure was shared with us by Episcopal Relief & Development, which they use in their internal program monitoring and evaluation. This measure included two questions: "Do you feel that you have things in common with other caregivers in your community?" and "Do you feel that other caregivers in your community care about you?" Caregivers rated their responses on a scale from strongly disagree to strongly agree, and the average score was used to assess overall community connectedness. The third variable was caregiver depressive symptoms, measured using the standardized 10-item version of the Center for Epidemiologic Studies Depression Scale (CESD-10). Another variable was the caregivers' level of financial worries (e.g., concerns about

buying food, paying monthly household expenses, and covering child-related costs). We developed this measure specifically for this study. Unlike other outcomes, which were assessed at both baseline and endline, financial worries were measured only at endline, as the measure was developed and added later in the evaluation. As a result, the analysis for this outcome compares scores between intervention versus control groups at endline but not changes over time. Finally, we compared the proportion of caregivers who reported receiving a referral for themselves or their children in the past year. Similar to the financial worries variable, this referral measure was included only in the endline survey.

Quantitative Data Analysis

We estimated the effects of the intervention at six-month follow-up on each outcome using mixed-effects (i.e., multilevel) regression models. Specifically, we included fixed effects that controlled for key sociodemographic factors (i.e., child's age, gender, primary caregiver age, whether the primary caregiver was the child's mother, and household wealth quintile), the study county, and, where available, the baseline measure of the outcome. An indicator for whether the follow-up interview was conducted by phone was also included in analyses. We included a random intercept at the village level to account for the clustered study design. All analyses were based on intention-to-treat (ITT) analysis, which minimizes bias in RCT analyses by including all participants in the trial analysis according to their original treatment assignment (intervention versus control group), even if some participants did not fully participate in the intervention or even dropped out. ITT analysis is considered the gold standard for interpreting RCTs because it provides a more realistic estimate of the treatment effect under real-world conditions, where adherence to interventions is often not perfect.

Ultimately for continuous outcomes, we report intervention effect sizes as standardized mean differences (β), scaled to the standard deviation (*SD*) of the outcome, to facilitate in the magnitude across different outcomes and ease of interpretation. In social science research, including in the field of parenting programs and ECD, effect sizes of 0.20 are considered small, 0.50 is moderate, and 0.80 or above is large (Cohen, 1988). Binary outcomes were reported as odds ratios (*OR*). Analyses were conducted in Stata. Statistical significance was set at $p < 0.05$.

RESULTS

Quantitative Impact Evaluation

Sociodemographic characteristics of the trial cohort at baseline

To briefly recap the sample characteristics at baseline, the trial included 595 primary caregivers of children under 18 months of age, with 310 caregivers in the intervention group and 285 in the control group (Table 5). The average age of the index child was 8.7 months (SD = 4.5), with no significant difference between study arms. Most primary caregivers were the child's mother (90.6%), with a significantly greater proportion of mothers in the control group compared to the intervention group (95.8% vs. 85.8%, respectively; p<0.01). Grandmothers were the next most common caregivers (7.6%), with more as the primary caregiver in the intervention group than the control group (11.6% vs. 3.2%; p<0.01). On average, five individuals lived in each household (SD = 1.8), and 77.5% of caregivers reported being currently in a partnered relationship (married or living with their partner). There were no significant differences by study group in household size or partnership status. In terms of caregiver age, most were between 25–34 years (45.2%) or 18–24 years (34.5%), with more caregivers 45+ years in the intervention group than the control group (9.0% vs 2.8%), likely due to the intervention's inclusion of grandmothers as primary caregivers. While the intervention allowed for the enrollment of grandmothers in this role, fewer households in the control group had a grandmother serving as the primary caregiver, which is why there was a smaller proportion of older caregivers in the control group. Educational attainment was similar across study arms. Overall, 43.4% of caregivers had completed secondary school, 41.7% had completed primary school, 14.2% had some primary schooling, and less than 1% had no formal education.

Table 5. Sample demographic characteristics at baseline by intervention arm.

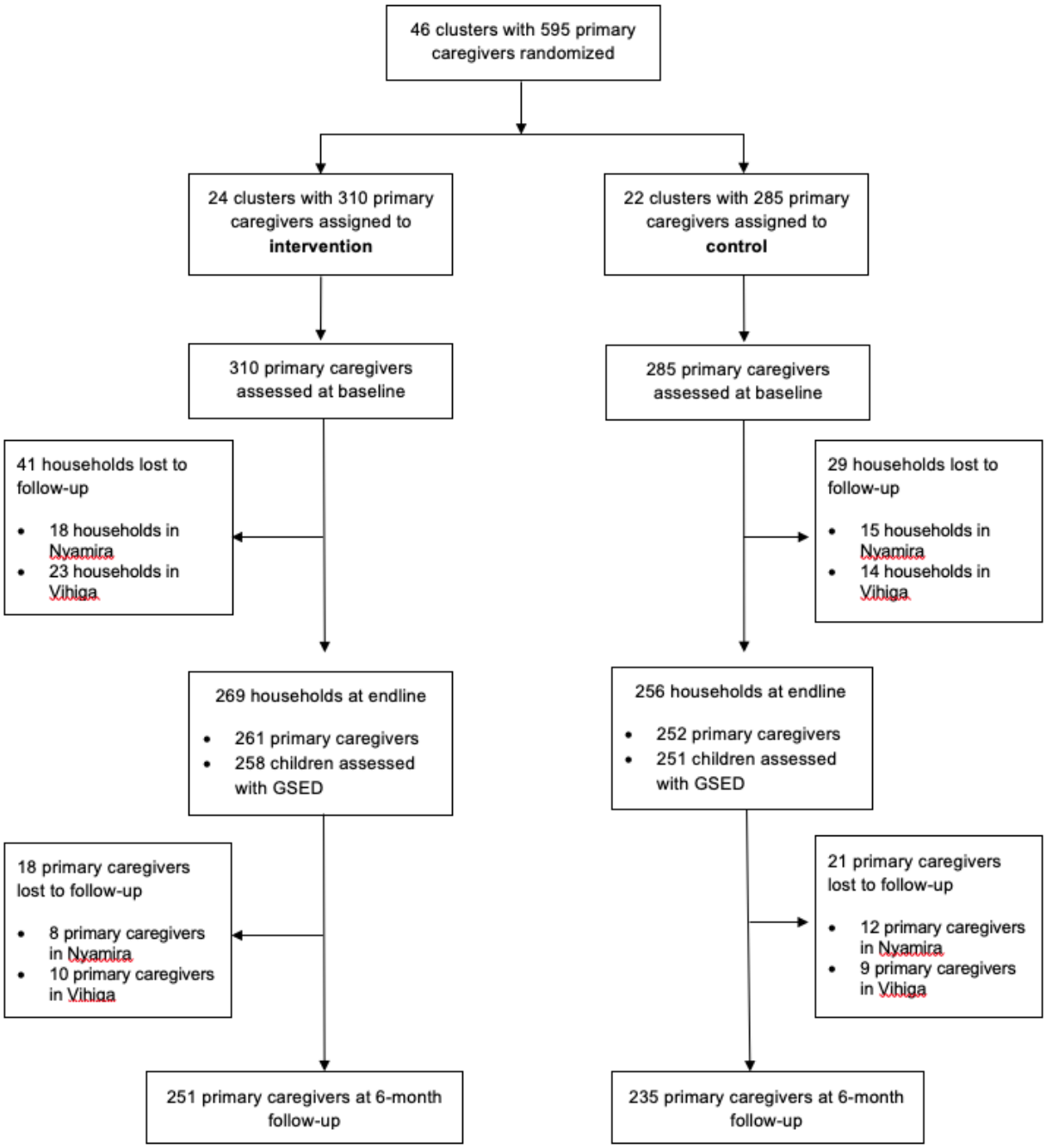
		Study Arm		
	Overall (N=595)	**Intervention (N=310)**	**Control (N=285)**	**P-value**
Child age in months, *mean (sd)*	8.7 (4.5)	8.9 (5.3)	8.6 (5.6)	0.58
Child sex, *n (%)*				
Male	271 (45.6%)	145 (46.8%)	126 (44.2%)	0.530
Female	324 (54.5%)	165 (53.2%)	159 (55.8%)	
Primary caregiver relation to child, *n (%)*				
Mother	539 (90.6%)	273 (85.8%)	266 (95.8%)	<0.01
Father	6 (1.0%)	4 (1.3%)	2 (0.7%)	0.47
Grandmother	45 (7.6%)	36 (11.6%)	9 (3.2%)	<0.01
Grandfather	2 (0.3%)	2 (0.7%)	0 (0.0%)	0.17
Aunt	2 (0.3%)	1 (0.3%)	1 (0.4%)	0.95
Other	1 (0.2%)	0 (0.0%)	1 (0.3%)	0.34
Number of individuals living in household, *mean (sd)*	5.02 (1.8)	4.9 (1.9)	5.2 (1.7)	0.09
Currently in partnered relationship, *n (%)*	461 (77.5%)	239 (77.1%)	222 (77.9%)	0.82
Primary caregiver age, *n (%)*				

18-24 years	205 (34.5%)	107 (34.5%)	98 (34.4%)	<0.01
25-34 years	269 (45.2%)	125 (40.3%)	144 (50.5%)	
35-44 years	85 (14.3%)	50 (16.1%)	35 (12.3%)	
45+ years	36 (6.1%)	28 (9.0%)	8 (2.8%)	
Primary caregiver education completion, *n (%)*				
No education	5 (0.8%)	3 (1.0%)	2 (0.7%)	0.32
Some primary school (incomplete)	84 (14.2%)	51 (16.5%)	22 (11.6%)	
Completed primary school	248 (41.7%)	129 (41.6%)	119 (41.8%)	
Completed secondary school	258 (43.4%)	127 (41.0%)	131 (46.0%)	
Household wealth quintile, *n (%)*				
Lowest	119 (20.0%)	67 (21.6%)	52 (18.3%)	0.66
Second	120 (20.2%)	63 (20.3%)	57 (20.0%)	
Middle	119 (20.0%)	65 (21.0%)	54 (19.0%)	
Fourth	119 (20.0%)	57 (18.4%)	62 (21.8%)	
Highest	118 (19.8%)	58 (18.7%)	60 (21.1%)	

Sample reassessed for endline and six-month data collection

Figure 1 shows the trial flow diagram and the sample that was reassessed at baseline and endline. As mentioned above in the Methods, out of the original 595 caregivers we reassessed 525 primary caregivers at endline (88% follow-up rate) and 486 at both endline and six-month follow-up (82% follow-up rate). Ultimately, this revisit rate was within our sample size calculation, which assumed an 80% follow-up rate.

Figure 1. Participant flow diagram for sample assessed at baseline, endline, and six-month follow-up in cluster-RCT.



Sustained participation and minimal risk of contamination

Of the 251 intervention participants assessed at follow-up, 107 (42.6%) reported meeting with their Caregiver Support and Learning Group (CSLG) at least once post-intervention (i.e., between endline and the 6-month follow-up).

Additionally, we descriptively examined the risk of contamination in the control group. At baseline, only 2.8% of control participants reported attending any parenting education or support groups (including CSLGs) in the prior 3 months. This remained consistently low in the control group at endline (8.7%) and follow-up (5.8%), reflecting only modest and relatively small changes over time. In contrast, participation among intervention participants in any parenting education or support groups in the prior 3 months rose substantially from 6.1% at baseline to 62.4% at endline, and 48.3% at follow-up, which is likely a reflection of their involvement in MTM CSLGs. Thus, the overall risk of contamination was deemed minimal for the control group.

Impact evaluation results

In this section, we present the six-month follow-up impact evaluation results by organizing outcomes into thematic areas (e.g., ECD, early learning, psychosocial wellbeing). As a reminder and noted in the Methods section, all quantitative outcome data are reported by the primary caregiver. This includes outcome data on male caregivers, which reflect the primary caregivers' reports of changes in male caregiver outcomes.

When describing intervention effects on each outcome, we report the unadjusted mean scores or proportions to note how these values changed between groups and over time; see Table 6 for all the descriptive statistics for each outcome by time and study arm. Also see Appendix 3 for visual plots of these same values of unadjusted means and percentages for each outcome – which includes all 3 time points of baseline, endline, and follow-up – by intervention versus control group. Note, these figures are intended to illustrate general trends over time and should be interpreted at intervention effects, as they are not based on statistical modeling or adjustment of covariates.

While the descriptive details are included in the appendices, below we present the intervention effect sizes for each outcome – SDs for continuous outcomes and ORs for binary outcomes. These effect sizes are derived from adjusted regression models that control for baseline scores and adjusts for various covariates (e.g., child age, caregiver education, household wealth). These adjusted effect sizes based on multilevel regression models are plotted in the figures below. Each point represents the standardized difference in the outcome between the intervention and control groups at six-month follow-up, with the extending lines representing the 95% confidence interval. Table 7 provides the specific values for these intervention effect sizes, which correspond to the values plotted in the figures.

ECD outcomes

Impacts of the intervention on ECD outcomes are illustrated in Figure 2. Broadly, there were minimal observed intervention impacts across ECD outcomes at follow-up. Although both the intervention and control group showed improvements in average CREDI-overall scores from baseline to follow-up ($M = 44.7$ to $M = 52.7$ for intervention, $M = 44.2$ to $M = 52.5$ for control), there was no significant difference between the two groups in CREDI-overall scores at follow-up

(β = 0.09, p = .447). Similar patterns were observed for the CREDI cognition, language, and motor subscales. The difference between intervention groups was marginally significant for the social-emotional subscale (β = 0.21, p = .056), indicating a small intervention effect sustained at six-month follow-up. The SDQ-overall likewise showed no significant difference in between intervention (M = 12.09) and control (M = 12.46) at follow-up (β = -0.09, p = .328). Similar patterns were observed for the SDQ prosocial, internalizing, and externalizing subscales.

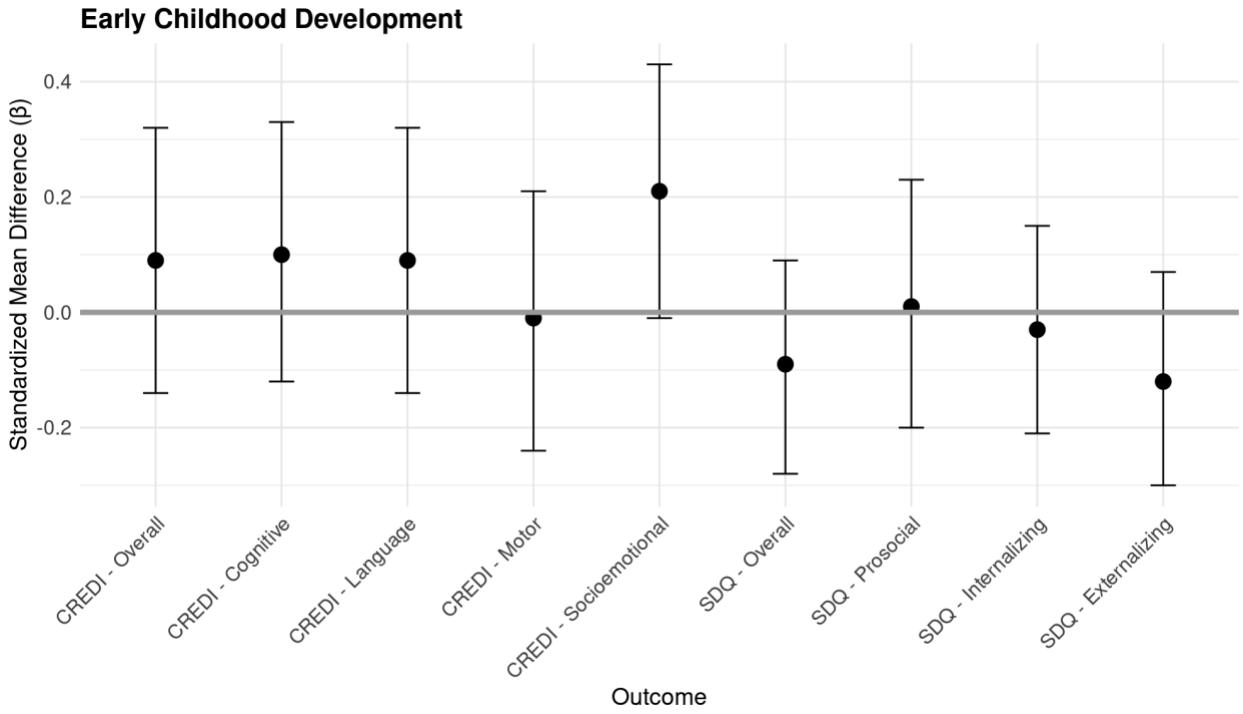


Figure 2. Intervention effects (β) on ECD at 6-mo follow-up.

Early learning

Figure 3 presents evidence of the modest-to-moderate intervention effects on early learning outcomes at follow-up. Primary caregiver stimulation increased between baseline and follow-up in both the intervention (M = 7.1 to M = 9.1) and control (M = 6.8 to M = 7.9) groups. However, this increase was larger in the intervention group, with an effect size of β = 0.39 SD difference between the intervention and control groups at follow-up (p < .001). Similar small but statistically significant intervention impacts were observed at follow-up for learning materials (β = 0.24, p = .021) and books (β = 0.28, p = .008), indicating sustained intervention impact in early learning outcomes through follow-up.

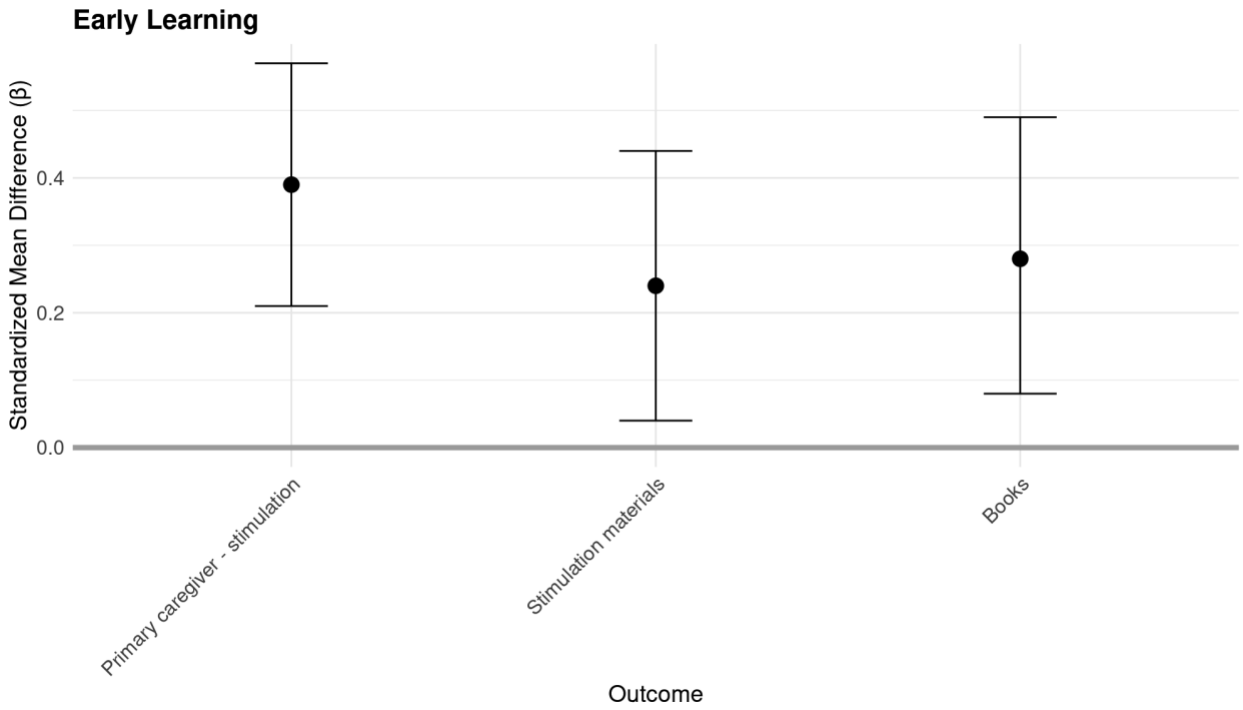


Figure 3. Intervention effects (β) on early learning at 6-mo follow-up.

Safety and security

Results for intervention effects on primary caregiver discipline practices and birth registration are presented in Figure 4. There were no significant differences between groups at follow-up in the use of any positive discipline strategies by primary caregivers ($OR = 0.51$, $p = .203$), though both control and intervention groups saw large increases (53% to 97% and 55% to 94%, respectively). However, the intervention had significant impacts on the use of harsh discipline practices. The use of any violent discipline (i.e., physical punishment or psychological aggression) increased in the control group from 58% to 88% but did not substantively change in the intervention group from 60% to 62%. This difference corresponded to 81% lower odds of using harsh discipline at follow-up in the intervention group compared to the control group ($OR = 0.19$, $p < .001$). Similar patterns were observed for physical punishment ($OR = 0.37$, $p < .001$) and psychological aggression ($OR = 0.29$, $p = .001$): whereas the control group showed marked increases in harsh punishment from baseline to follow-up, the intervention group remained relatively stable. The proportion of children with birth registration increased over time in both groups—from 35% to 53% in the intervention group and from 31% to 49% in the control group—but the difference between groups at follow-up was not statistically significant ($p = .295$).

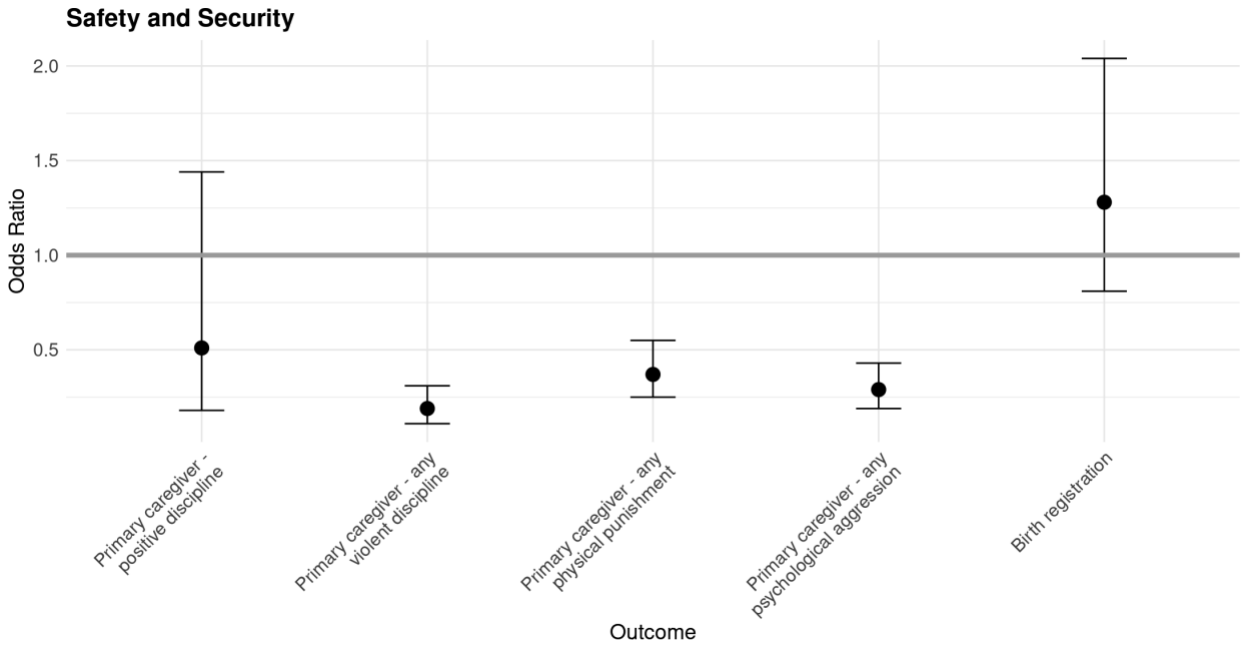


Figure 4. Intervention effects (OR) on safety and security at 6-mo follow-up.

Nutrition, health, household food and financial security

We examined intervention effects on a range of nutrition, health, and economic outcomes, presented in Figures 6 and 7; generally, findings showed some intervention effect on nutrition and health outcomes. Dietary diversity improved on average in the intervention group from baseline to follow-up (from $M = 3.4$ to 4.3) whereas it remained relatively stable in the control group (from $M = 3.5$ to 3.7), translating to an intervention impact at follow-up of $\beta = 0.34$ *SD* ($p = .001$). The proportion of children aged 6–23 months who received a minimum acceptable diet increased in the intervention group from baseline to follow-up (29% to 48%) but remained stable in the control group (from 28% to 31%); at follow-up, the intervention group had over two times the odds of children having a minimum acceptable diet relative to the control group ($OR = 2.23$, $p = .009$). There was also an intervention impact at follow-up on caregiver-reported child illness, such that the intervention group had 39% lower odds of reporting child illness relative to the control (OR = .61, $p = .023$); from baseline to follow-up, the intervention group reported fewer child illnesses (78% to 68%) whereas the control group remained relatively stable (72% to 76%). There were no significant differences in appropriate care-seeking behaviors or receipt of referrals at follow-up. Additionally, there were no significant intervention impacts on kitchen garden access or financial security outcomes (e.g., income, savings) at follow-up.

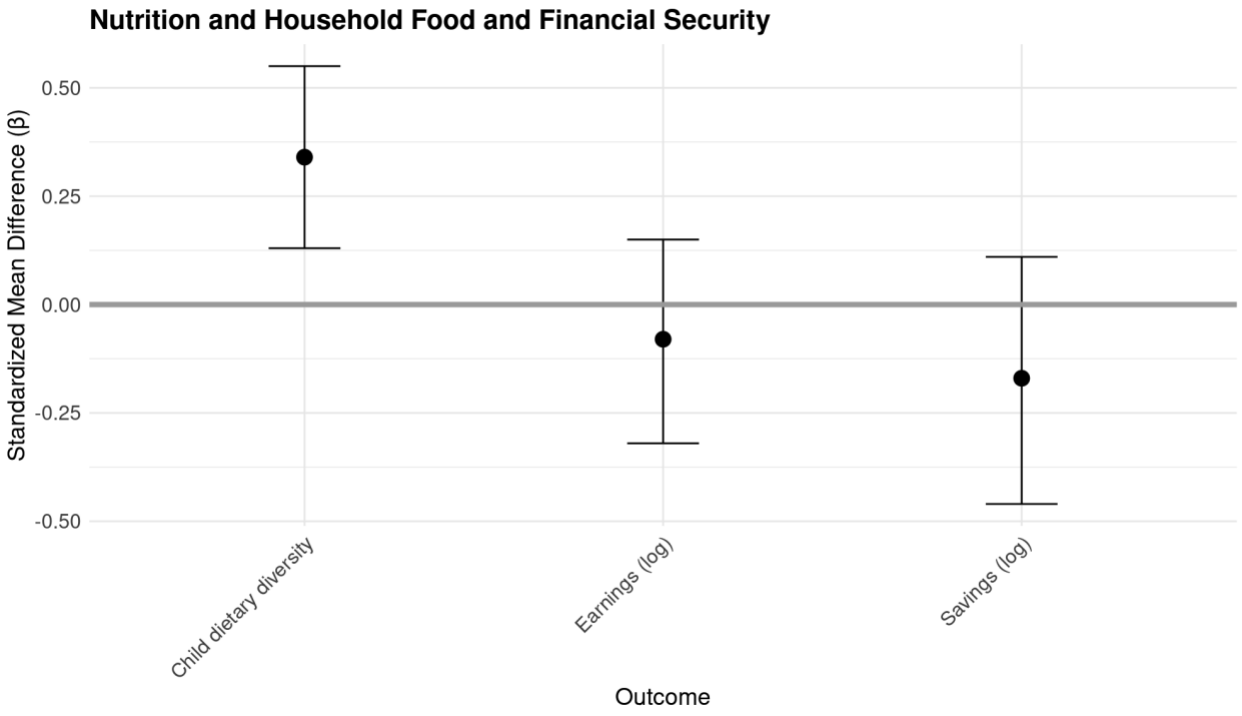


Figure 5. Intervention effects (β) on nutrition, household food, and financial security at 6-mo follow-up.

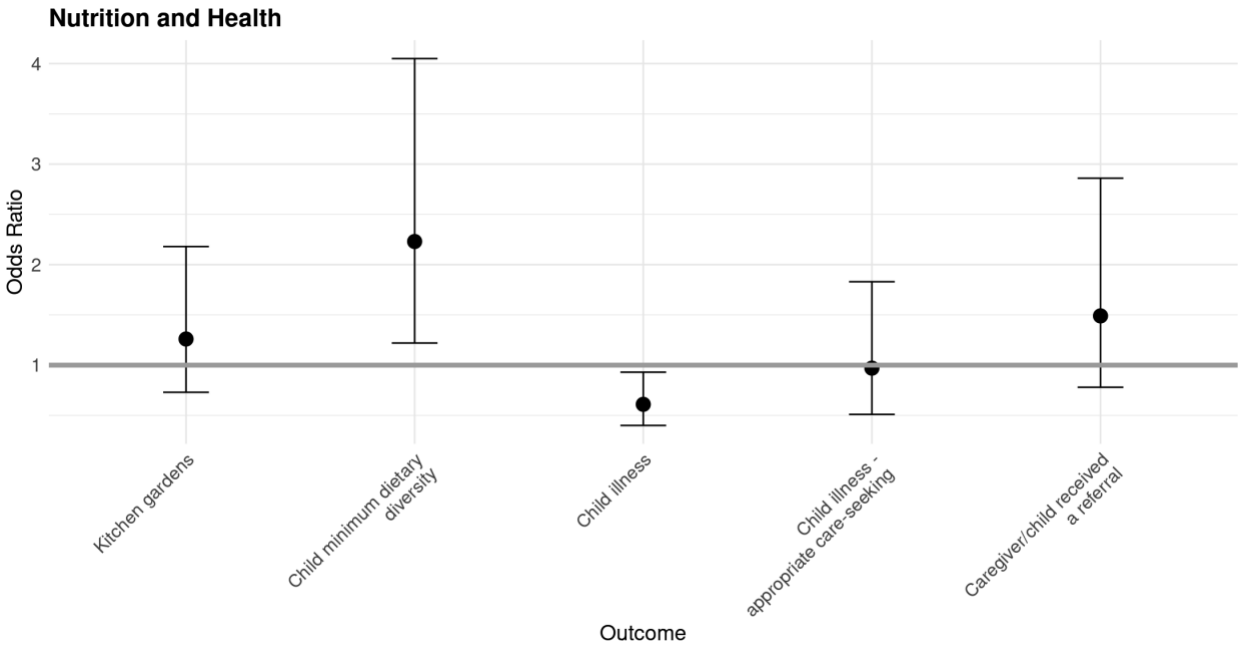


Figure 6. Intervention effects (ORs) on nutrition and health at 6-mo follow-up.

Psychosocial wellbeing

Intervention impacts on psychosocial wellbeing outcomes are presented in Figure 7. Generally, there were sustained intervention impacts at follow-up for psychosocial wellbeing outcomes. Primary caregivers in the intervention group had improvements in average social support from baseline to follow-up (M = 3.4 to 3.6) whereas the control group showed no change (M = 3.4 to 3.3), corresponding to an intervention effect at follow-up of β = 0.42 SD (p < .001). Similarly, community connectedness also improved on average in the intervention group (M = 3.0 to 3.2) but was stable in the control group (M = 3.0 to 3.0), corresponding to an intervention effect at follow-up of β = 0.31 SD (p = .002). Although there was a small increase in parenting stress from baseline to follow-up in the intervention group (M = 33.9 to 34.8), the increase was even larger in the control group (M = 32.8 to 35.8), indicating the intervention may have provided a buffering effect on parental stress (β = -0.21, p = .016). Likewise, depression scores decreased slightly in the intervention group (M = 9.3 to 8.7) while increasing in the control group (M = 8.2 to 9.1), resulting in a protective intervention effect at follow-up (β = -0.17, p = .045). Financial worries were not significantly different between groups at follow-up.

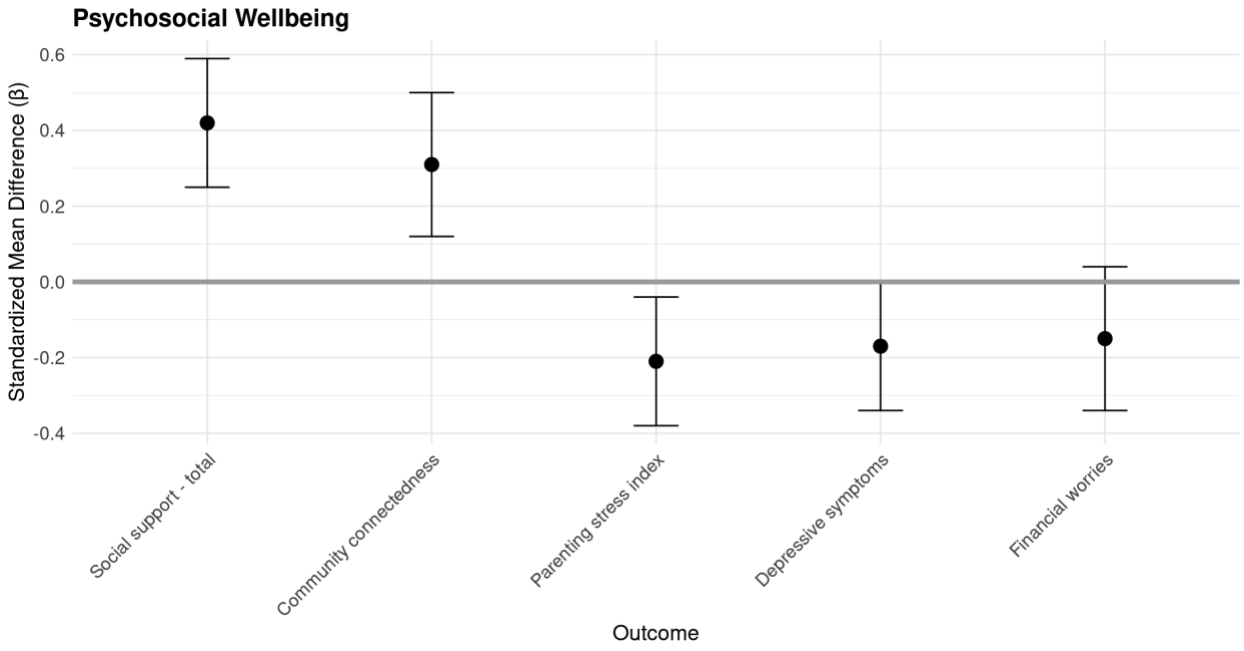


Figure 7. Intervention effects (β) on psychosocial wellbeing at 6-mo follow-up.

The intervention had significant effects on primary caregivers' reports of IPV (Figure 8). Reports of any IPV from baseline to follow-up in the intervention group (48% to 35%), yet remained relatively consistent in the control group (47% to 43%). This effect corresponded to an odds ratio of 0.51 (p = .028). Similarly, emotional IPV was significantly lower in the intervention compared

to control group at follow-up (29% vs. 37% respectively, *OR* = 0.51, *p* = .023), as was physical IPV (8% vs. 13%, *OR* = 0.45, *p* = .047), and economic IPV (18% vs. 28%, *OR* = 0.53, *p* = .017).

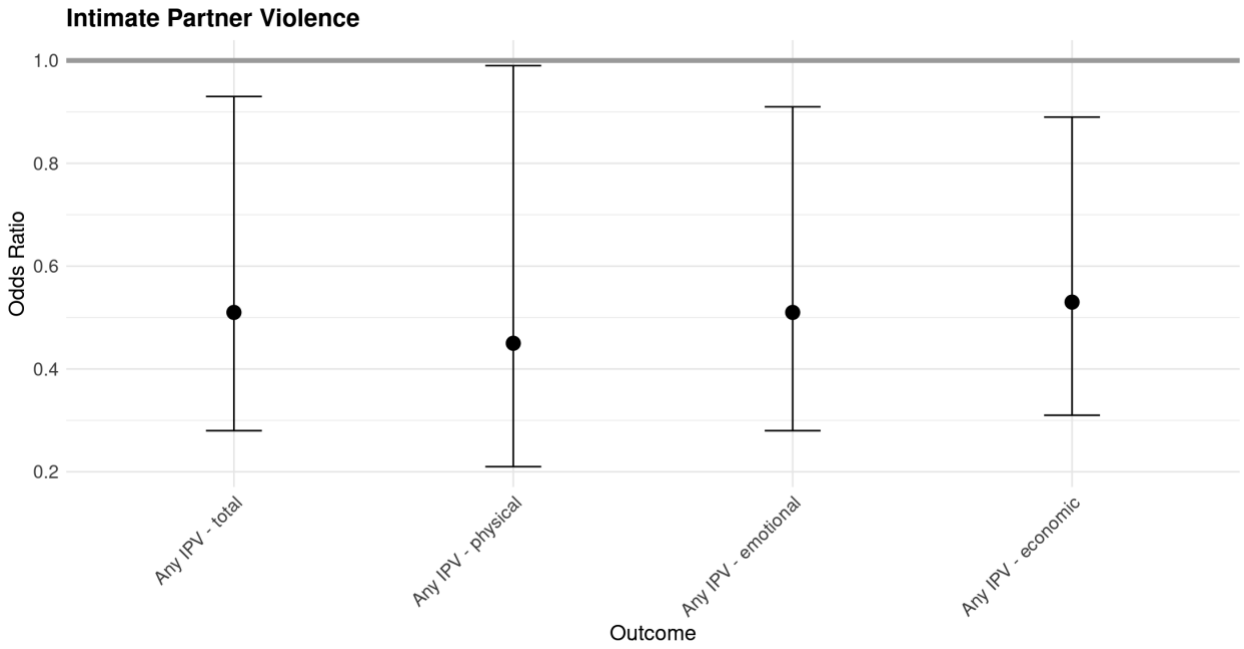


Figure 8. Intervention effects (ORs) on intimate partner violence victimization at 6-mo follow-up.

Male caregiver outcomes

We additionally assessed intervention effects at follow-up on male caregiver practices, which are presented in Figures 9 and 10 and tended to show favorable intervention effects. Male caregiver stimulation increased on average more in the intervention group (*M* = 4.8 to 6.4) than in the control group (*M* = 4.2 to 5.1) from baseline to follow-up, with an effect size of β = 0.34 *SD* (*p* < 0.001). Violent discipline by male caregivers increased in the control group from baseline to follow-up (30% to 34%) but remained stable in the intervention group (32% to 34%), resulting in a significant difference between groups at follow-up (*OR* = 0.63, *p* = .039). A similar trend was observed for male caregivers' use of psychological aggression (30% and 21% for control and intervention groups respectively at follow-up, *OR* = 0.50, *p* = .007). Physical punishment by male caregivers followed a similar trajectory (34% and 26% for control and intervention groups respectively at follow-up, *OR* = 0.65) but was marginally significant (*p* = .097). There was no significant difference in male caregivers' use of positive discipline between groups at follow-up (intervention: 36% to 78%; control: 30% to 72%). Male caregivers' involvement in household chores increased slightly in the intervention group (from *M* = 2.1 to 2.4) but declined in the control group (from 2.2 to 1.3), with a significant effect size of β = 0.52 *SD* (*p* < .001).

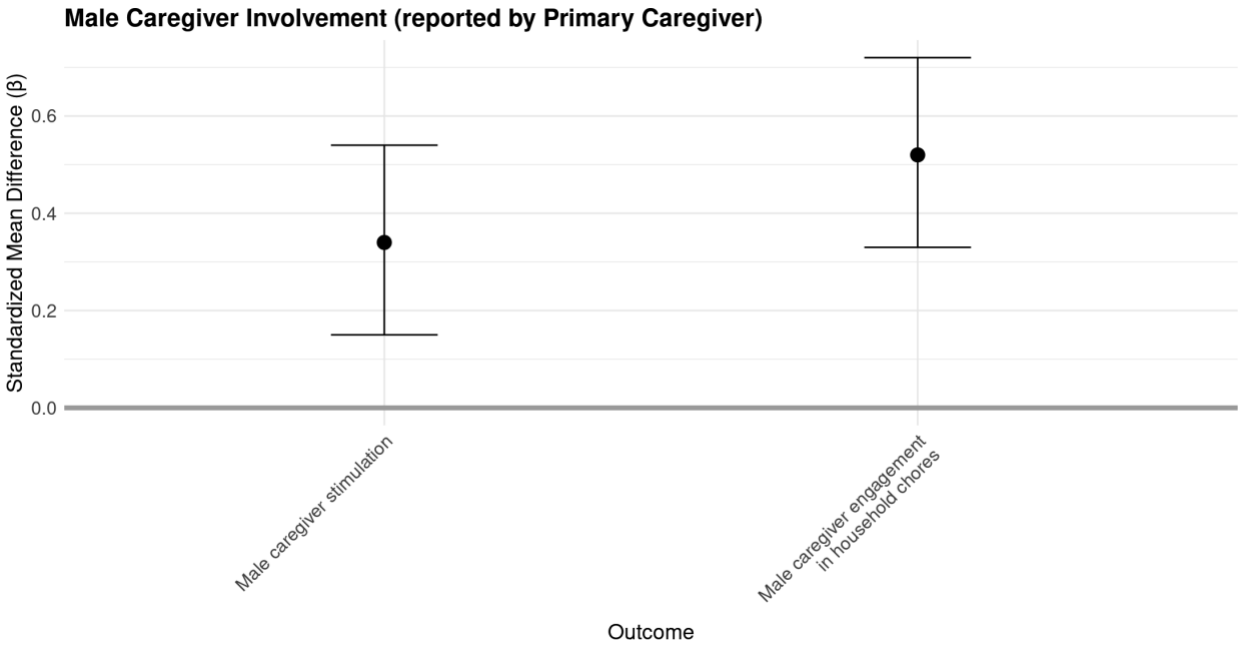


Figure 9. Intervention effects (β) on male caregiver involvement at 6-mo follow-up.

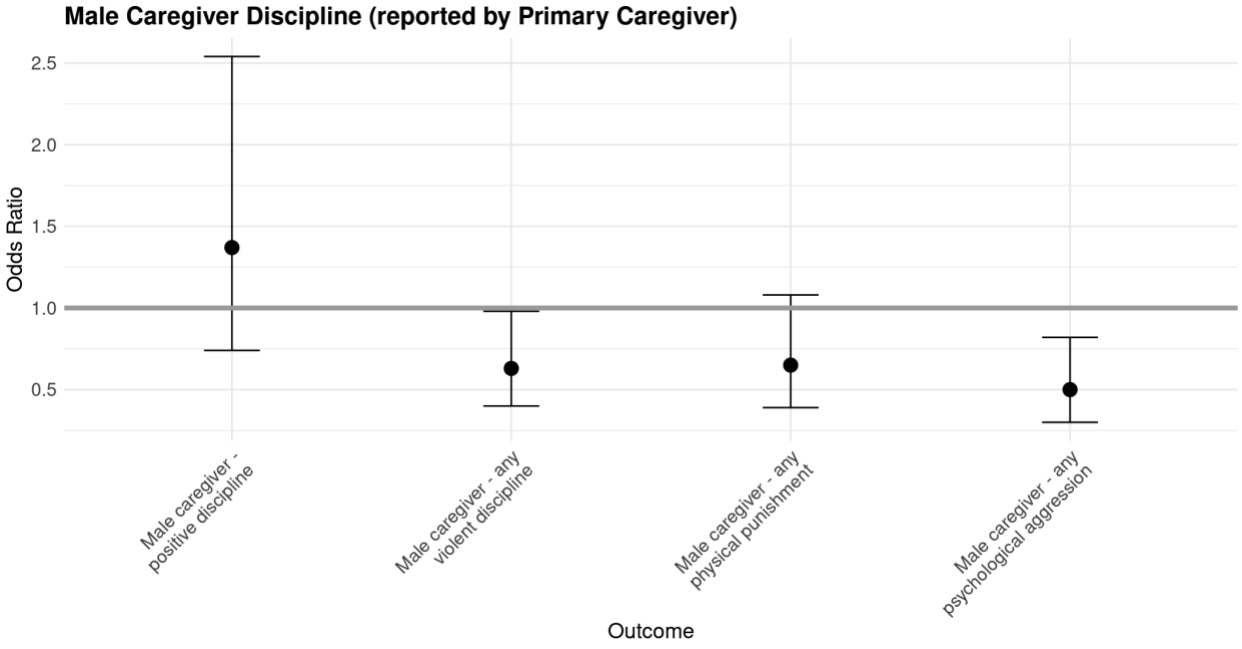


Figure 10. Intervention effects (OR) on male caregiver discipline at 6-mo follow-up.

Table 6. Descriptive statistics of outcomes, overall sample.

	Baseline			Endline			6-Month Follow-up		
	Control n=285	Intervention n=310	p-value	Control n=256	Intervention n=269	p-value	Control n=235	Intervention n=251	p-value
CREDI Overall score	44.2 (3.9)	44.7 (3.8)	0.103	51.8 (1.5)	52.1 (1.5)	**0.045**	52.5 (1.1)	52.7 (1.1)	0.294
CREDI Cognition score	47.0 (2.3)	47.2 (2.2)	0.109	50.7 (1.0)	50.9 (1.0)	**0.037**	51.1 (0.8)	51.3 (0.8)	0.324
CREDI Language score	47.3 (1.7)	47.5 (1.7)	0.145	51.5 (1.2)	51.7 (1.2)	0.053	52.1 (1.0)	52.2 (1.0)	0.330
CREDI Motor score	46.3 (2.6)	46.6 (2.6)	0.117	51.5 (1.2)	51.7 (1.2)	0.110	51.9 (0.9)	52.0 (1.0)	0.696
CREDI Social-emotional score	46.6 (2.6)	47.0 (2.4)	0.132	51.4 (1.1)	51.7 (1.1)	**0.013**	52.0 (0.9)	52.0 (0.8)	0.053
Primary caregiver stimulation index score (11 item)	6.8 (2.6)	7.1 (2.5)	0.167	8.8 (2.1)	9.6 (2.4)	**<0.001**	7.9 (3.1)	9.1 (2.8)	**<0.001**
Number of learning materials	2.2 (1.7)	2.5 (1.7)	**0.010**	3.9 (1.0)	4.3 (0.9)	**<0.001**	3.6 (1.1)	4.0 (1.2)	**0.001**
Number of books in household	0.5 (1.5)	0.7 (1.5)	0.172	1.4 (1.8)	2.0 (2.0)	**<0.001**	1.2 (1.6)	1.7 (2.3)	**0.002**
Primary caregiver use of any positive discipline	152 (53%)	172 (55%)	0.599	245 (97%)	247 (95%)	0.139	227 (97%)	236 (94%)	0.182
Primary caregiver use of any violent discipline	164 (58%)	186 (60%)	0.543	222 (88%)	144 (55%)	**<0.001**	208 (88%)	156 (62%)	**<0.001**
Primary caregiver use of any physical punishment	140 (49%)	156 (50%)	0.770	200 (79%)	126 (48%)	**<0.001**	177 (75%)	135 (54%)	**<0.001**
Primary caregiver use of any psychological aggression	130 (46%)	142 (46%)	0.962	173 (69%)	102 (39%)	**<0.001**	172 (73%)	118 (47%)	**<0.001**
Birth registration	87 (31%)	109 (35%)	0.253	144 (57%)	164 (63%)	0.188	116 (49%)	133 (53%)	0.424
Kitchen garden in household	206 (72%)	210 (68%)	0.228	184 (73%)	221 (85%)	**0.001**	168 (72%)	181 (72%)	0.879
Child dietary diversity score (24 hr)	3.5 (1.6)	3.4 (1.5)	0.336	3.9 (1.4)	4.4 (1.4)	**<0.001**	3.7 (1.5)	4.3 (1.6)	**<0.001**
Children aged 6+ months who receive a minimum dietary diversity (MDD) in past 24 hours	46 (28%)	55 (29%)	0.806	81 (33%)	113 (46%)	**0.003**	69 (31%)	115 (48%)	**<0.001**
Child experienced any illness (diarrhea, cough, or fever) in last 2 weeks	206 (72%)	241 (78%)	0.124	143 (57%)	143 (56%)	0.838	178 (76%)	168 (68%)	0.035
Appropriate care (hospital, clinic, CHV) sought for any child illness in past 2 weeks	100 (49%)	150 (62%)	**0.004**	64 (45%)	77 (54%)	0.124	71 (40%)	81 (48%)	0.119
Caregiver/child received a referral in past 6 months				35 (14%)	78 (30%)	**0.000**	32 (13%)	43 (17%)	0.284

Income in past month (KSH)	1057.4 (2524.7)	1289.1 (3735.8)	0.380	3219.1 (5176.9)	3037.5 (4890.9)	0.738	4076.2 (5056.1)	3876.1 (6576.9)	0.787
Total amount currently in savings (i.e Bank, SACCO etc) (KSH)	1892.8 (18053.2)	769.1 (3316.8)	0.282	2202.4 (7393.8)	2831.5 (5055.9)	0.260	2815.1 (11096.7)	2592.5 (4434.3)	0.769
Money accessed in credit in past month (KSH)	21047.7 (112862.7)	35104.2 (222838.7)	0.608	9441.0 (23839.3)	5243.2 (11469.3)	0.099	16576.6 (31353.1)	5369.5 (8681.1)	**0.001**
Primary caregiver overall social support total score	3.4 (0.7)	3.4 (0.6)	0.769	3.3 (0.8)	3.6 (0.7)	**<0.001**	3.3 (0.8)	3.6 (0.7)	**<0.001**
Community connectedness mean score	3.0 (1.0)	3.0 (0.9)	0.753	2.9 (1.1)	3.3 (1.1)	**<0.001**	3.0 (1.2)	3.2 (1.1)	**0.002**
Primary caregiver - any IPV victimization (physical, emotional or economic)	110 (47%)	120 (48%)	0.858	94 (47%)	67 (31%)	**0.001**	81 (43%)	73 (35%)	0.122
Primary caregiver - any physical IPV victimization	37 (16%)	47 (19%)	0.396	30 (15%)	20 (9%)	0.078	25 (13%)	17 (8%)	0.105
Primary caregiver - any emotional IPV victimization	95 (40%)	105 (42%)	0.781	78 (39%)	52 (24%)	**0.001**	70 (37%)	61 (29%)	0.110
Primary caregiver - any economic IPV victimization	76 (32%)	71 (28%)	0.317	59 (29%)	34 (16%)	**<0.001**	53 (28%)	37 (18%)	**0.016**
Primary caregiver parenting stress total score	32.8 (6.2)	33.9 (7.0)	0.065	33.4 (7.5)	31.0 (7.3)	**<0.001**	35.8 (5.4)	34.8 (5.8)	0.070
Primary caregiver depression total score	8.2 (5.9)	9.3 (6.7)	0.050	9.3 (7.2)	7.8 (6.3)	**0.010**	9.1 (6.9)	8.7 (6.6)	0.554
Primary caregiver - financial worries in past month				12.9 (5.1)	11.7 (5.4)	**0.011**	12.3 (5.2)	12.0 (5.1)	0.625
Father stimulation index score (11 item)	4.2 (3.4)	4.8 (3.5)	0.075	5.9 (4.0)	7.6 (4.0)	**<0.001**	5.1 (4.0)	6.4 (4.3)	**0.001**
Father use of any positive discipline	67 (30%)	82 (36%)	0.216	176 (80%)	174 (81%)	0.884	142 (72%)	159 (78%)	0.205
Father use of any violent discipline	66 (30%)	72 (32%)	0.717	109 (50%)	58 (27%)	**<0.001**	85 (43%)	70 (34%)	0.064
Father use of any psychological aggression	45 (20%)	42 (18%)	0.586	73 (33%)	39 (18%)	**<0.001**	60 (30%)	44 (21%)	**0.040**
Father use of any physical punishment	51 (23%)	61 (27%)	0.383	81 (37%)	41 (19%)	**<0.001**	67 (34%)	54 (26%)	0.094
Father involvement in household chores subscale score	2.2 (2.3)	2.1 (2.3)	0.773	1.1 (1.8)	2.2 (2.5)	**<0.001**	1.3 (1.9)	2.4 (2.4)	**<0.001**

Note. Descriptive statistics reflect individuals who had baseline, endline, and follow-up data on each outcome.

Table 7. Intervention effects on outcomes at 6-mo follow-up based on adjusted regression models and in the overall sample.

Outcome	*n*	Est. Type	Est.	95% CI LB	95% CI UB	*p*
Early Childhood Development						
CREDI overall	273	β	0.09	-0.14	0.32	0.447
CREDI cognition	273	β	0.10	-0.12	0.33	0.359
CREDI language	273	β	0.09	-0.14	0.32	0.430
CREDI motor	273	β	-0.01	-0.24	0.21	0.922
CREDI social-emotional	273	β	0.21	-0.01	0.43	0.056
SDQ overall	507	β	-0.09	-0.28	0.09	0.328
SDQ prosocial	507	β	0.01	-0.20	0.23	0.909
SDQ internalizing	507	β	-0.03	-0.21	0.15	0.710
SDQ externalizing	507	β	-0.12	-0.30	0.07	0.221
Early Learning						
Primary caregiver stimulation	506	β	**0.39**	**0.21**	**0.57**	**<.001**
No. of learning materials	503	β	**0.24**	**0.04**	**0.44**	**0.021**
No. of books	499	β	**0.28**	**0.08**	**0.49**	**0.008**
Child Safety and Security						
Primary caregiver positive discipline	507	OR	0.51	0.18	1.44	0.203
Primary caregiver violent discipline (any)	507	OR	**0.19**	**0.11**	**0.31**	**<.001**
Primary caregiver physical punishment	507	OR	**0.37**	**0.25**	**0.55**	**<.001**
Primary caregiver psychological aggression	507	OR	**0.29**	**0.19**	**0.43**	**0.001**
Birth registration	505	OR	1.28	0.81	2.04	0.295
Caregiver Psychosocial Wellbeing						
Social support	507	β	**0.42**	**0.25**	**0.59**	**<.001**
Community connectedness	503	β	**0.31**	**0.12**	**0.50**	**0.002**
Parenting stress	507	β	**-0.21**	**-0.38**	**-0.04**	**0.016**
Primary caregiver depressive symptoms	507	β	**-0.17**	**-0.34**	**0.00**	**0.045**
Primary caregiver financial worries	507	β	-0.15	-0.34	0.04	0.123
Any IPV	376	OR	**0.51**	**0.28**	**0.93**	**0.028**
Physical IPV	376	OR	**0.45**	**0.21**	**0.99**	**0.047**

Emotional IPV	376	OR	**0.51**	**0.28**	**0.91**	**0.023**
Economic IPV	376	OR	**0.53**	**0.31**	**0.89**	**0.017**
Nutrition, Health, Household Food, and Financial Security						
Kitchen garden	507	OR	1.26	0.73	2.18	0.414
Child dietary diversity (past 24 hr)	306	β	**0.34**	**0.13**	**0.55**	**0.001**
Child minimum diet diversity	284	OR	**2.23**	**1.22**	**4.05**	**0.009**
Child illness (past 2 weeks)	501	OR	**0.61**	**0.40**	**0.93**	**0.023**
Sought appropriate care for illness	283	OR	0.97	0.51	1.83	0.925
Received referral	507	OR	1.49	0.78	2.86	0.229
Income in past month (KSH, log)	273	β	-0.08	-0.32	0.15	0.474
Current savings (KSH, log)	220	β	-0.17	-0.46	0.11	0.232
Male Caregiver Outcomes						
Male caregiver stimulation	352	β	**0.34**	**0.15**	**0.54**	**<.001**
Male caregiver positive discipline	353	OR	1.37	0.74	2.54	0.317
Male caregiver violent discipline (any)	353	OR	**0.63**	**0.4**	**0.98**	**0.039**
Male caregiver physical punishment	353	OR	0.65	0.39	1.08	0.097
Male caregiver psychological aggression	353	OR	**0.50**	**0.30**	**0.82**	**0.007**
Male caregiver household chores	413	β	**0.52**	**0.33**	**0.72**	**<.001**

Note. **Bold: $p < .05$**

Variation in program effectiveness by county

Given the county-level differences in program effects observed at endline, we also examined whether outcomes at the 6-month follow-up varied by county. Descriptive statistics for outcome distributions in Nyamira and Vihiga are presented in Appendices 1 and 2, respectively. Table 8 presents adjusted regression results for intervention effects at the 6-month follow-up, stratified by county, with Nyamira shown in the left panel and Vihiga in the right. Figures 11-19 provide visual representations of the county-level effect size comparisons shown in the table.

At the 6-month follow-up, county-level differences were less pronounced than at endline, particularly for ECD. No statistically significant differences were found by county for any ECD outcomes. However, consistent with endline findings, Nyamira continued to show significantly larger effects for early learning outcomes – specifically, primary caregiver stimulation and the presence of books in the home – compared to Vihiga. Similarly, the increase in referral cases was again significant in Nyamira but not in Vihiga. Male involvement in household chores also showed a significant positive effect at 6-month follow-up only in Nyamira, mirroring endline results. These county-level patterns suggest a continuation for at least some of the differential impacts but not all previously observed at endline.

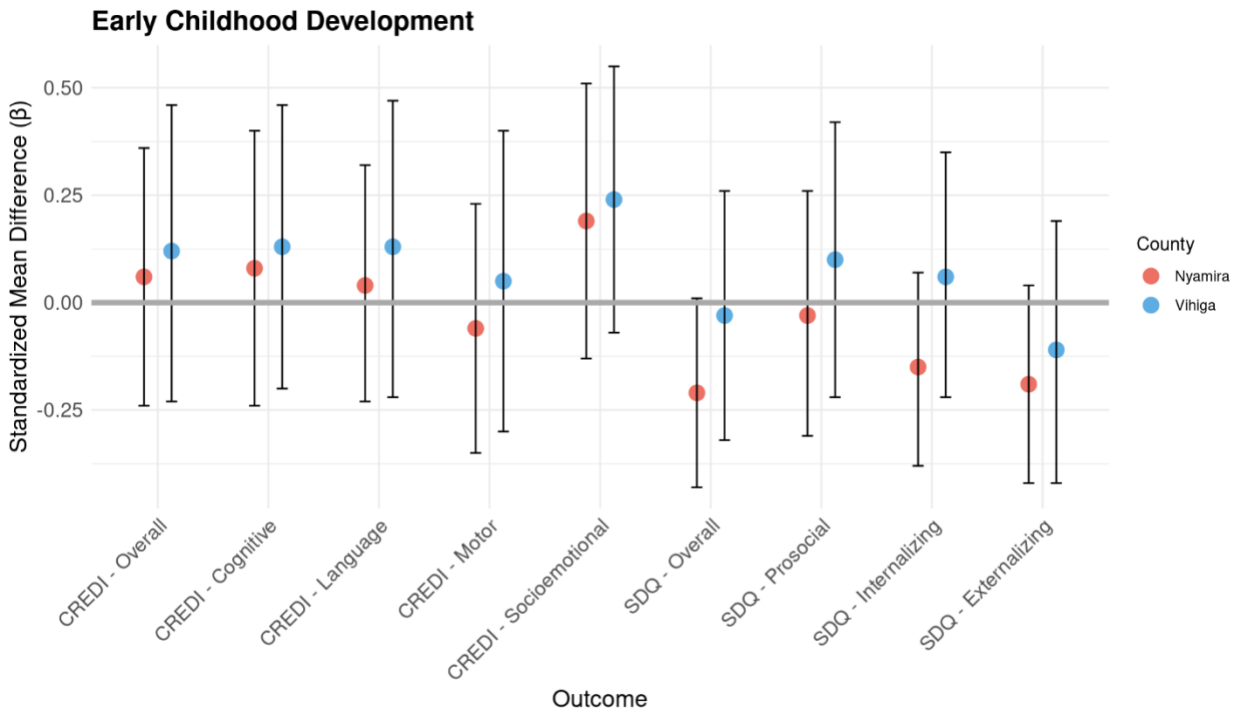


Figure 11. Intervention effects on early childhood development at 6-month follow-up, by county.

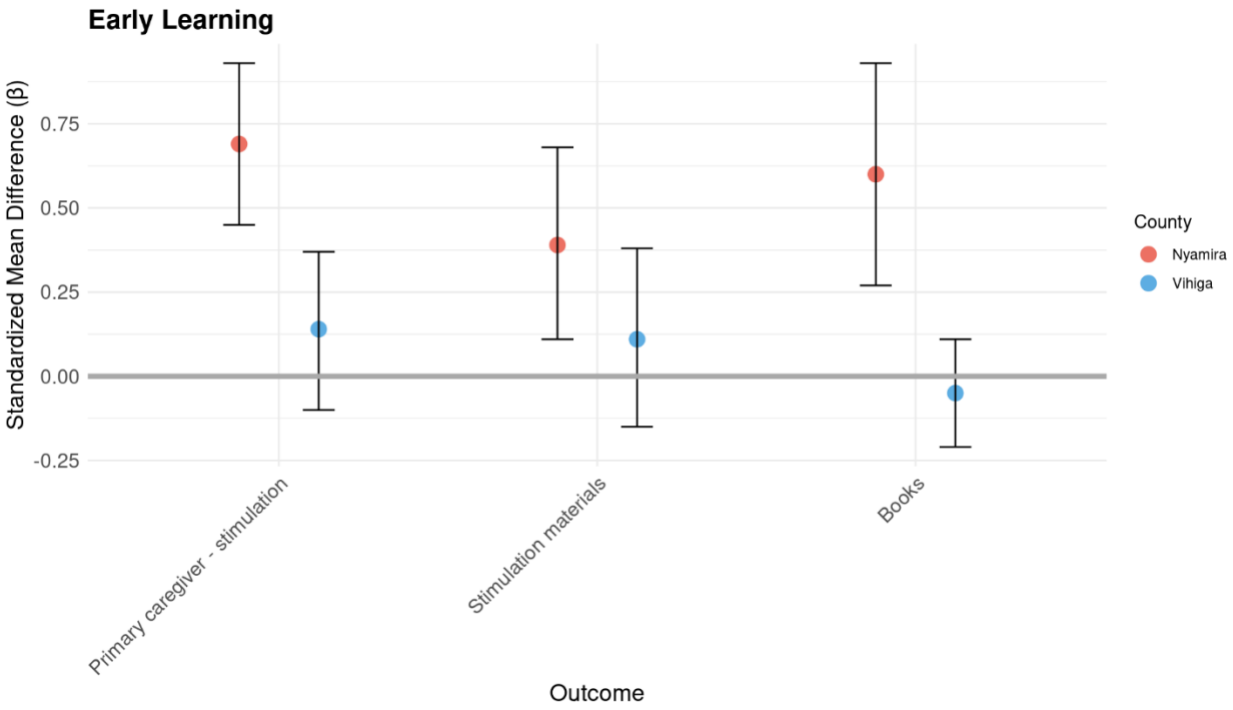


Figure 12. Intervention effects on early learning at 6-month follow-up, by county.

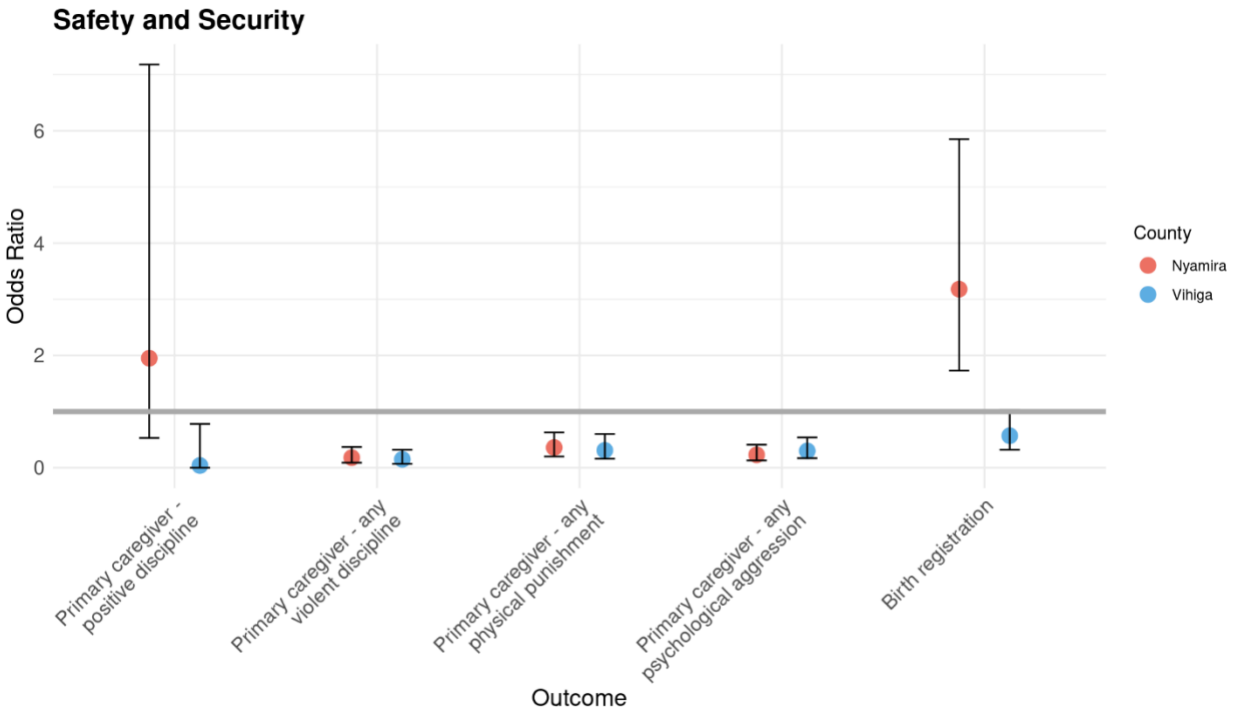


Figure 13. Intervention effects on safety and security at 6-month follow-up, by county.

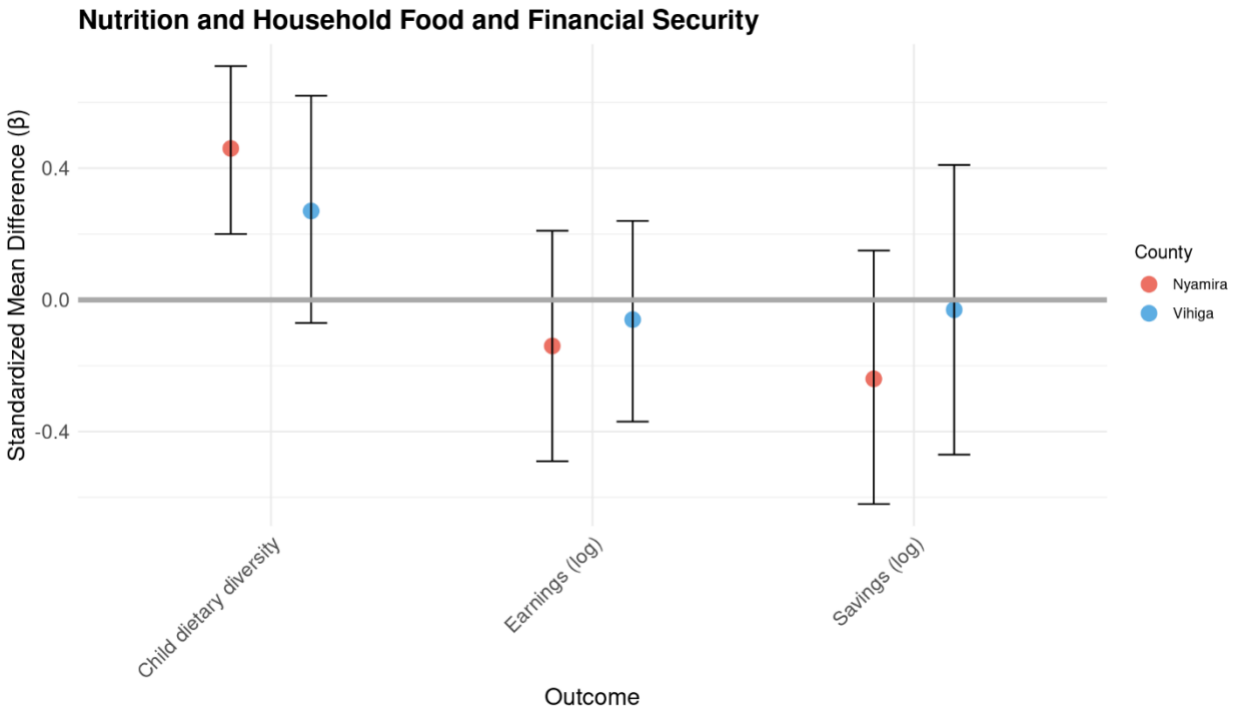


Figure 14. Intervention effects (β) on nutrition, household food, and financial security at 6-mo follow-up, by county

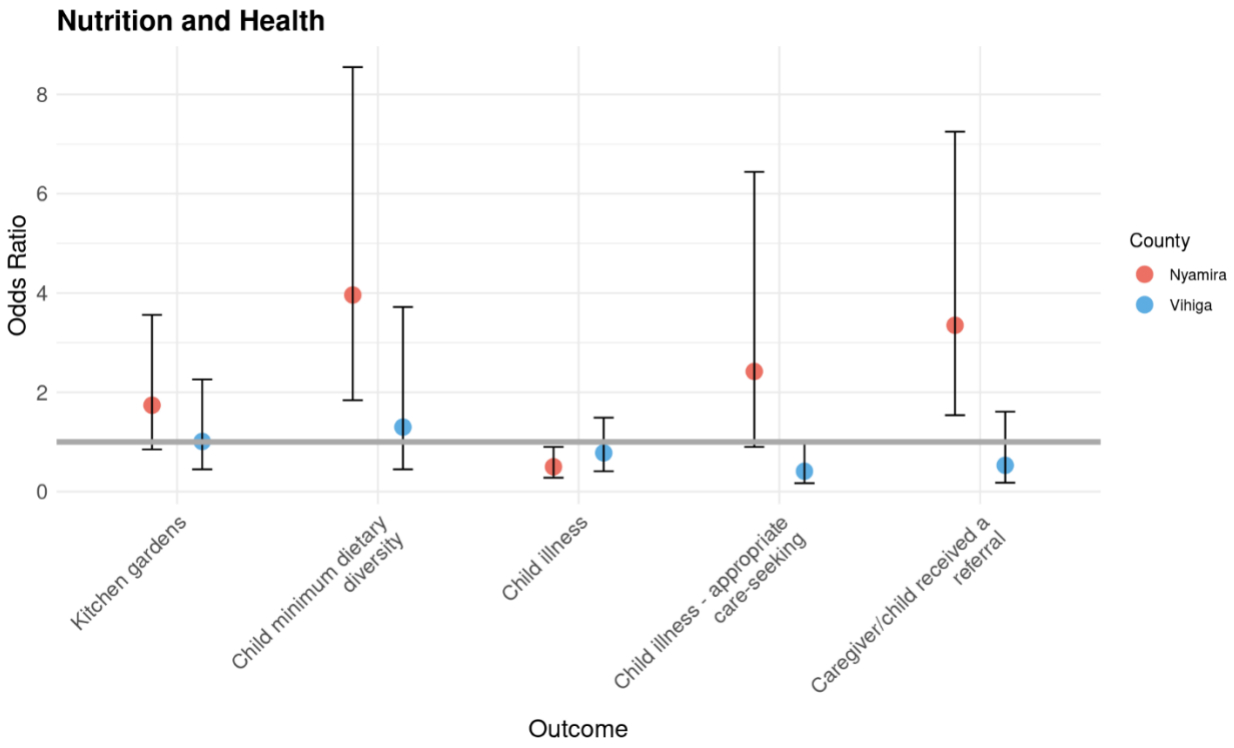


Figure 15. Intervention effects (OR) on nutrition and health at 6-mo follow-up, by county.

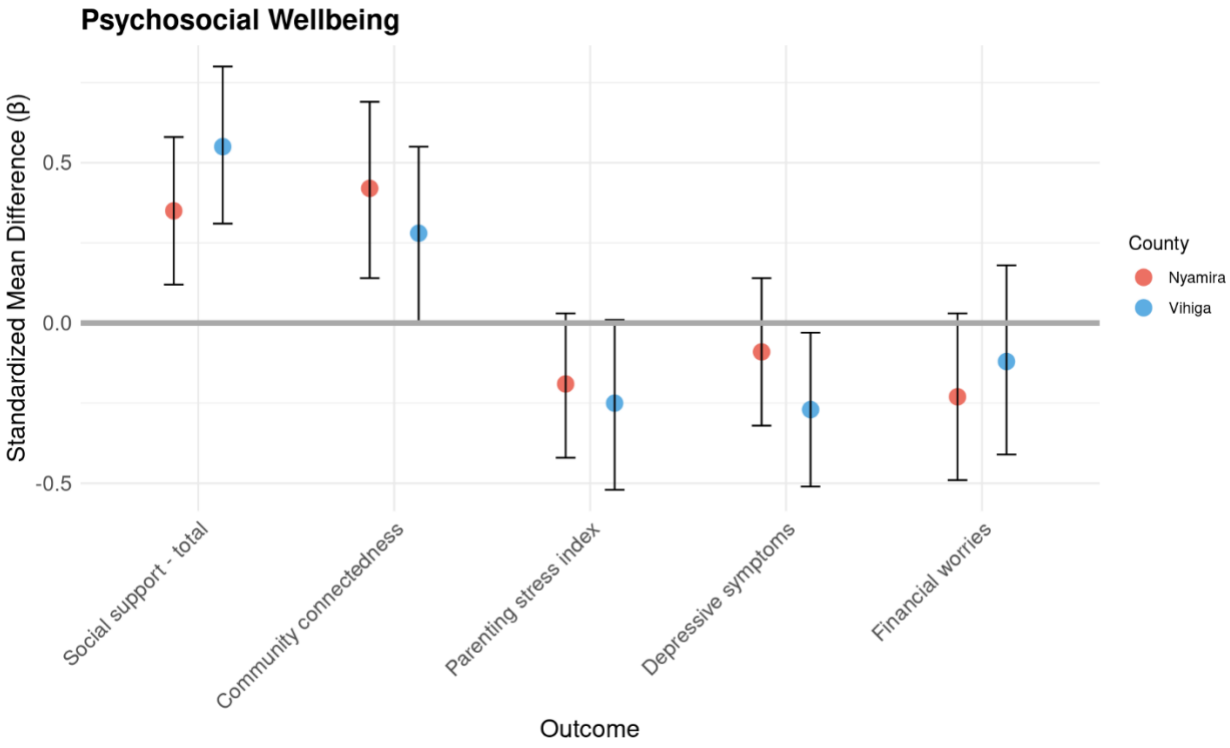


Figure 16. Intervention effects (β) on psychosocial wellbeing at 6-mo follow-up, by county.

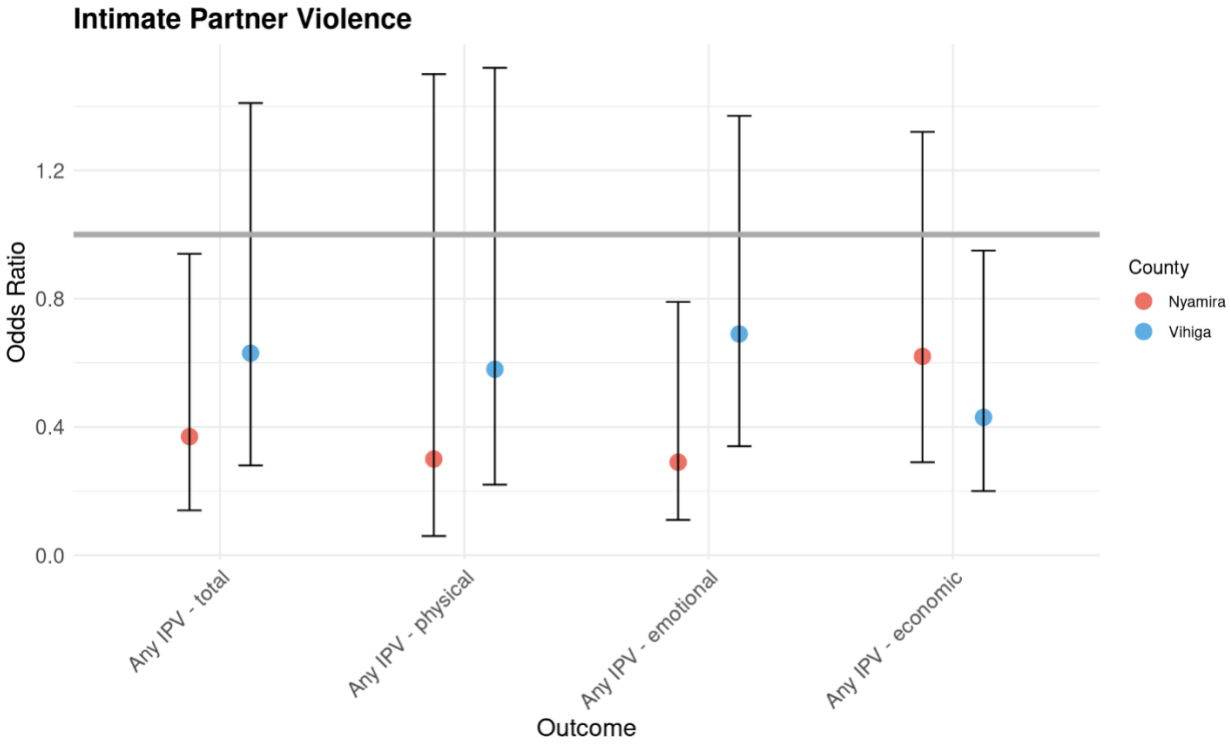


Figure 17. Intervention effects (ORs) on intimate partner violence victimization at 6-mo follow-up, by county

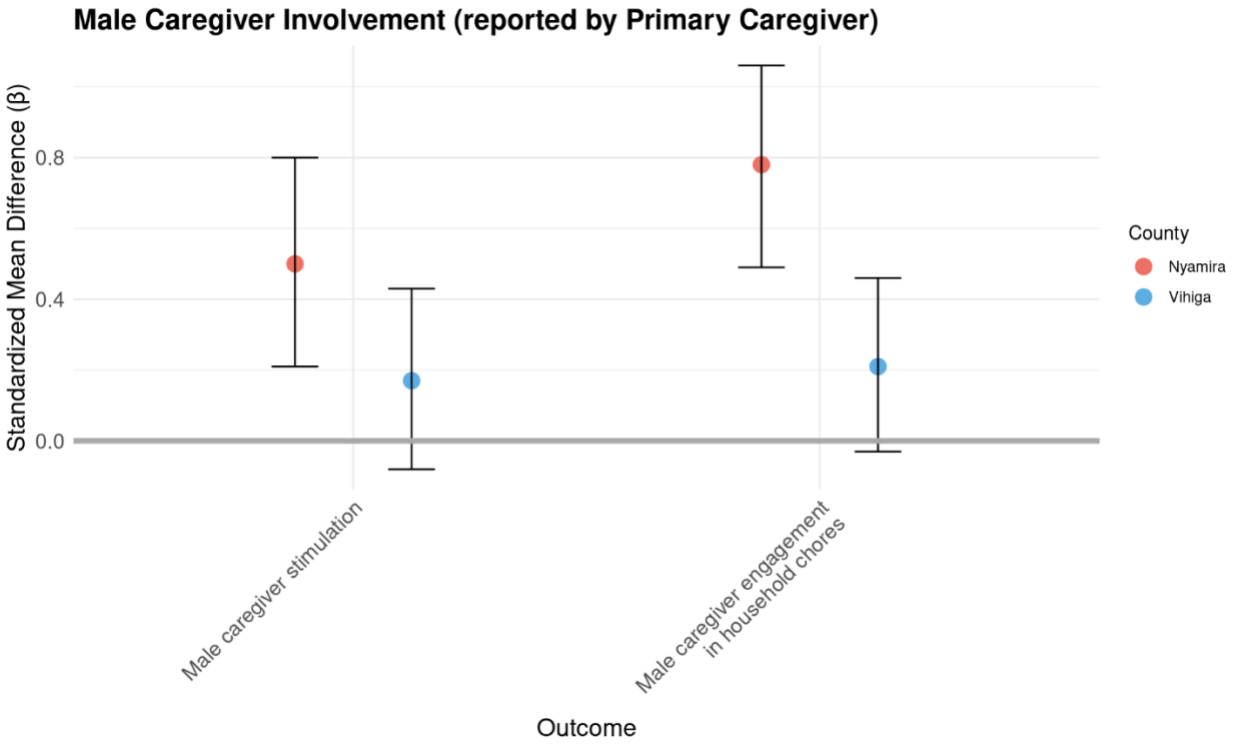


Figure 18. Intervention effects (β) on male caregiver involvement at 6-mo follow-up, by county

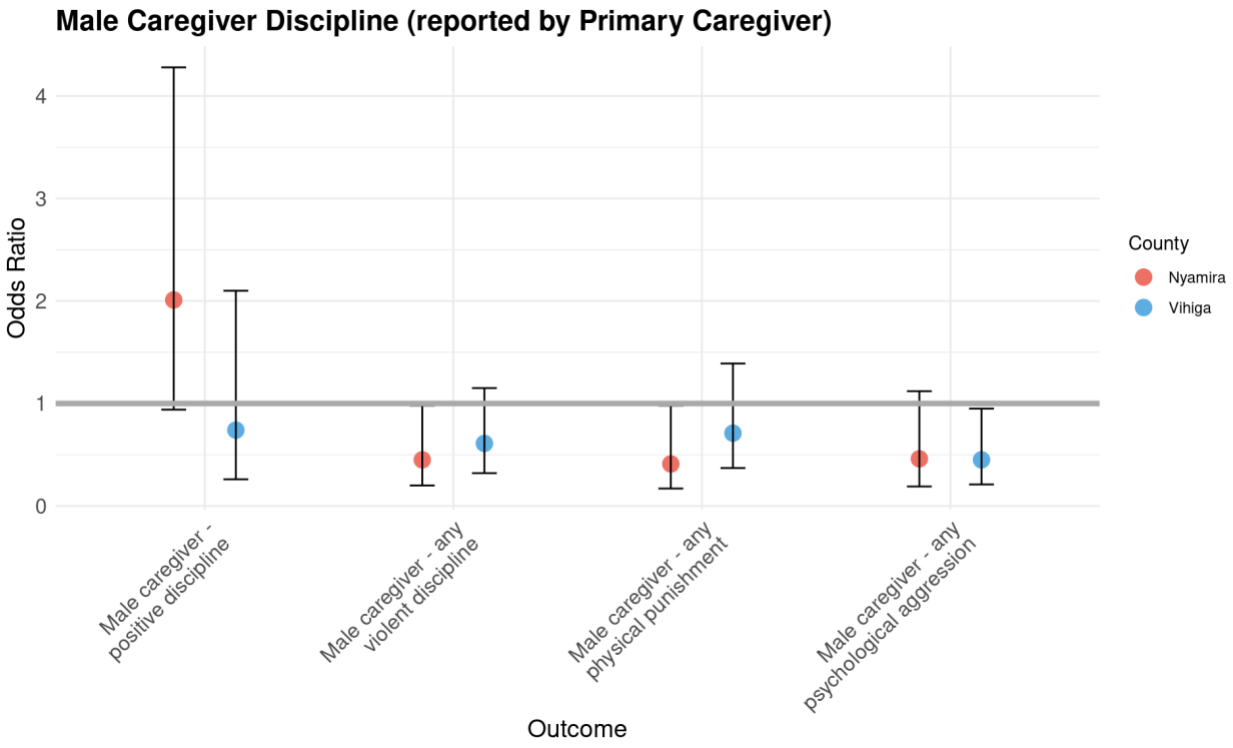


Figure 19. Intervention effects (β) on male caregiver discipline at 6-mo follow-up, by county.

Table 8. Intervention effects on outcomes at 6-mo follow-up based on adjusted regression models, stratified by county.

	Nyamira						Vihiga					
Outcome	***n***	**Est. Type**	**Est.**	**95% CI LB**	**95% CI UB**	***p***	***n***	**Est. Type**	**Est.**	**95% CI LB**	**95% CI UB**	***p***
Early Childhood Development												
CREDI overall	131	β	0.06	-0.24	0.36	0.688	142	β	0.12	-0.23	0.46	0.501
CREDI cognition	131	β	0.08	-0.24	0.40	0.609	142	β	0.13	-0.20	0.46	0.436
CREDI language	131	β	0.04	-0.23	0.32	0.755	142	β	0.13	-0.22	0.47	0.471
CREDI motor	131	β	-0.06	-0.35	0.23	0.697	142	β	0.05	-0.30	0.40	0.772
CREDI social-emotional	131	β	0.19	-0.13	0.51	0.245	142	β	0.24	-0.07	0.55	0.136
SDQ overall	254	β	-0.21	-0.43	0.01	0.067	253	β	-0.03	-0.32	0.26	0.833
SDQ prosocial	254	β	-0.03	-0.31	0.26	0.856	253	β	0.10	-0.22	0.42	0.549
SDQ internalizing	254	β	-0.15	-0.38	0.07	0.186	253	β	0.06	-0.22	0.35	0.653
SDQ externalizing	254	β	-0.19	-0.42	0.04	0.109	253	β	-0.11	-0.42	0.19	0.464
Early Learning												
Primary caregiver stimulation	254	β	**0.69**	**0.45**	**0.93**	**<.001**	252	β	0.14	-0.10	0.37	0.266
No. of learning materials	251	β	**0.39**	**0.11**	**0.68**	**0.007**	252	β	0.11	-0.15	0.38	0.394
No. of books	247	β	**0.60**	**0.27**	**0.93**	**<.001**	252	β	-0.05	-0.21	0.11	0.543
Child Safety and Security												
Primary caregiver positive discipline	218	OR	1.95	0.53	7.18	0.314	253	OR	**0.04**	**0.00**	**0.78**	**0.003**
Primary caregiver violent discipline (any)	254	OR	**0.18**	**0.09**	**0.37**	**<.001**	253	OR	**0.15**	**0.07**	**0.32**	**<.001**
Primary caregiver physical punishment	254	OR	**0.36**	**0.20**	**0.63**	**<.001**	253	OR	**0.31**	**0.16**	**0.60**	**<.001**
Primary caregiver psychological aggression	254	OR	**0.23**	**0.13**	**0.41**	**<.001**	253	OR	**0.30**	**0.17**	**0.54**	**<.001**
Birth registration	253	OR	**3.18**	**1.73**	**5.85**	**<.001**	252	OR	0.57	0.32	1.03	0.061
Caregiver Psychosocial Wellbeing												
Social support	254	β	**0.35**	**0.12**	**0.58**	**0.003**	253	β	**0.55**	**0.31**	**0.80**	**<.001**
Community connectedness	251	β	**0.42**	**0.14**	**0.69**	**0.003**	252	β	**0.28**	**0.00**	**0.55**	**0.048**
Parenting stress	254	β	-0.19	-0.42	0.03	0.094	253	β	-0.25	-0.52	0.01	0.062

Primary caregiver depressive symptoms	254	β	-0.09	-0.32	0.14	0.455	253	β	**-0.27**	**-0.51**	**-0.03**	**0.030**
Primary caregiver financial worries	254	β	-0.23	-0.49	0.03	0.083	253	β	-0.12	-0.41	0.18	0.442
Any IPV	197	OR	**0.37**	**0.14**	**0.94**	**0.038**	179	OR	0.63	0.28	1.41	0.259
Physical IPV	197	OR	0.30	0.06	1.50	0.143	179	OR	0.58	0.22	1.52	0.267
Emotional IPV	197	OR	**0.29**	**0.11**	**0.79**	**0.016**	179	OR	0.69	0.34	1.37	0.287
Economic IPV	197	OR	0.62	0.29	1.32	0.213	179	OR	**0.43**	**0.20**	**0.95**	**0.036**
Nutrition, Health, Household Food, and Financial Security												
Kitchen garden	254	OR	1.74	0.85	3.56	0.132	253	OR	1.01	0.45	2.26	0.989
Child dietary diversity (past 24 hr)	160	β	**0.46**	**0.20**	**0.71**	**<.001**	146	β	0.27	-0.07	0.62	0.122
Child minimum diet diversity	148	OR	**3.96**	**1.84**	**8.55**	**<.001**	136	OR	1.30	0.45	3.72	0.625
Child illness (past 2 weeks)	251	OR	**0.50**	**0.28**	**0.90**	**0.020**	250	OR	0.78	0.41	1.49	0.455
Sought appropriate care for illness	131	OR	2.42	0.90	6.44	0.078	152	OR	**0.41**	**0.17**	**0.99**	**0.047**
Received referral	254	OR	**3.35**	**1.54**	**7.25**	**0.002**	253	OR	0.53	0.18	1.61	0.265
Income in past month (KSH, log)	148	β	-0.14	-0.49	0.21	0.429	125	β	-0.06	-0.37	0.24	0.679
Current savings (KSH, log)	128	β	-0.24	-0.62	0.15	0.224	92	β	-0.03	-0.47	0.41	0.897
Male Caregiver Outcomes												
Male caregiver stimulation	174	β	**0.50**	**0.21**	**0.80**	**0.001**	178	β	0.17	-0.08	0.43	0.183
Male caregiver positive discipline	176	OR	2.01	0.94	4.28	0.072	177	OR	0.74	0.26	2.10	0.577
Male caregiver violent discipline (any)	176	OR	**0.45**	**0.20**	**0.98**	**0.045**	177	OR	0.61	0.32	1.15	0.127
Male caregiver physical punishment	176	OR	**0.41**	**0.17**	**0.98**	**0.046**	177	OR	0.71	0.37	1.39	0.324
Male caregiver psychological aggression	176	OR	0.46	0.19	1.12	0.086	177	OR	**0.45**	**0.21**	**0.95**	**0.037**
Male caregiver household chores	210	β	**0.78**	**0.49**	**1.06**	**<.001**	203	β	0.21	-0.03	0.46	0.085

Note. **Bold: $p < .05$**

Comparison of intervention effects on various outcomes at endline vs 6-month follow-up

Finally, to summarize these results, we present visualizations that display post-intervention (endline) effects alongside 6-month follow-up outcomes, allowing comparison of whether effects were sustained or diminished over time. Figures 20-28 provide visual representations of the effect size comparisons between endline and the 6-month follow-up.

At six months post-intervention, many parenting and family environment outcomes remained strong, including improvements in caregiver stimulation, book use, and male caregiver involvement. Sustained reductions were also apparent in in harsh discipline practices and intimate partner violence. Positive effects on child diet diversity, caregiver nutrition knowledge, health service use, and community connectedness were also maintained. In contract, effects on overall ECD outcomes and specific domains (cognitive, language, motor, socioemotional) were evident at endline but tended to diminish by the 6-month follow-up, with estimates closer to zero. The only domain that approached statistical significance was socioemotional development ($\beta = 0.21$ SD, $p = 0.056$). Parenting stress, depressive symptoms, and financial stress all faded out at the 6-month mark. Positive discipline practices among male caregivers also weakened over time.

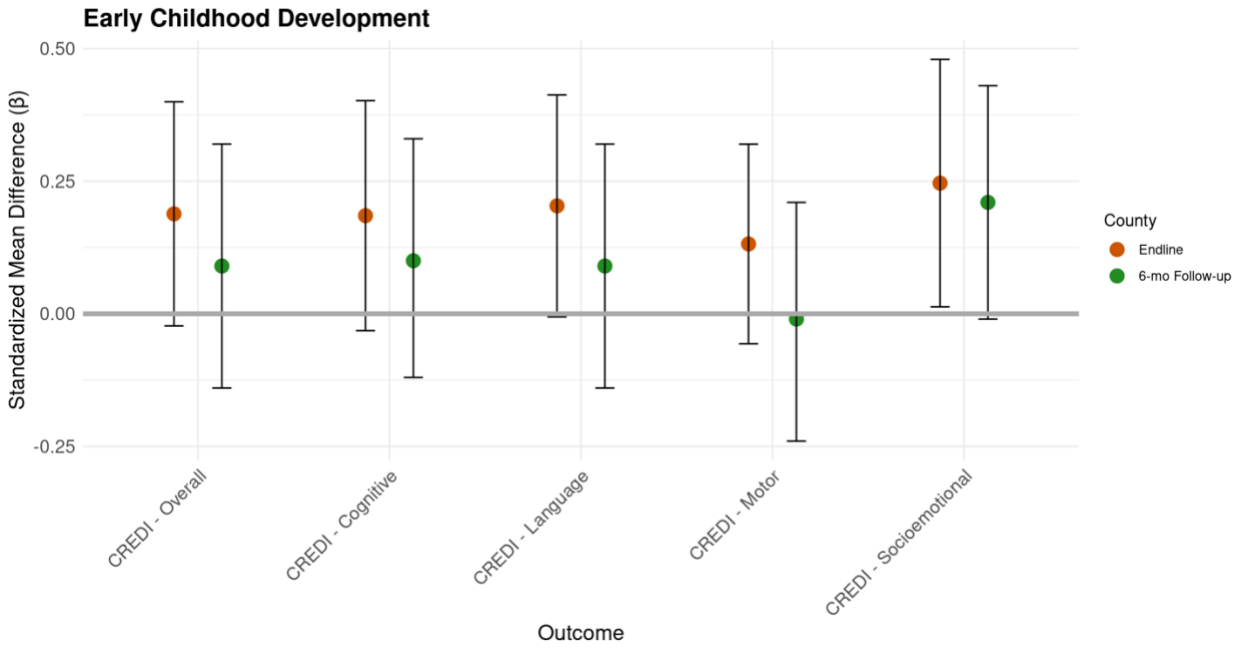


Figure 20. Intervention effects (β) on early childhood development at endline and 6-month follow-up

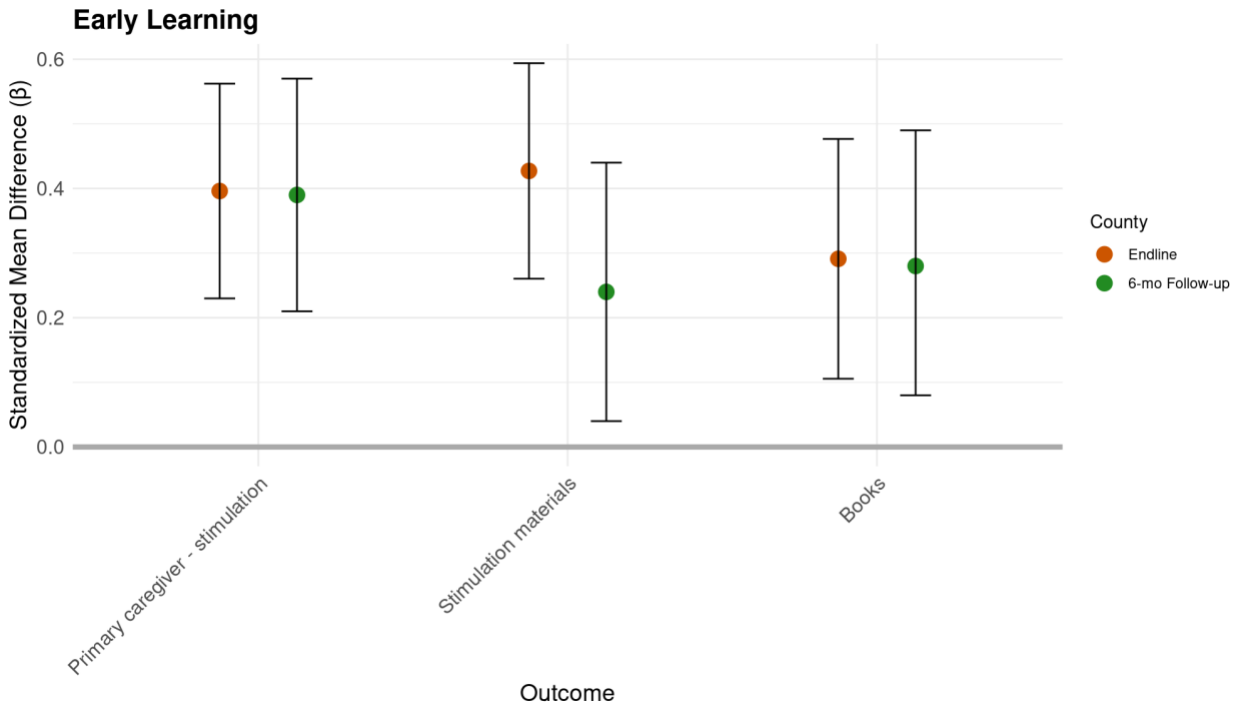


Figure 21. Intervention effects (β) on early learning at endline and 6-month follow-up

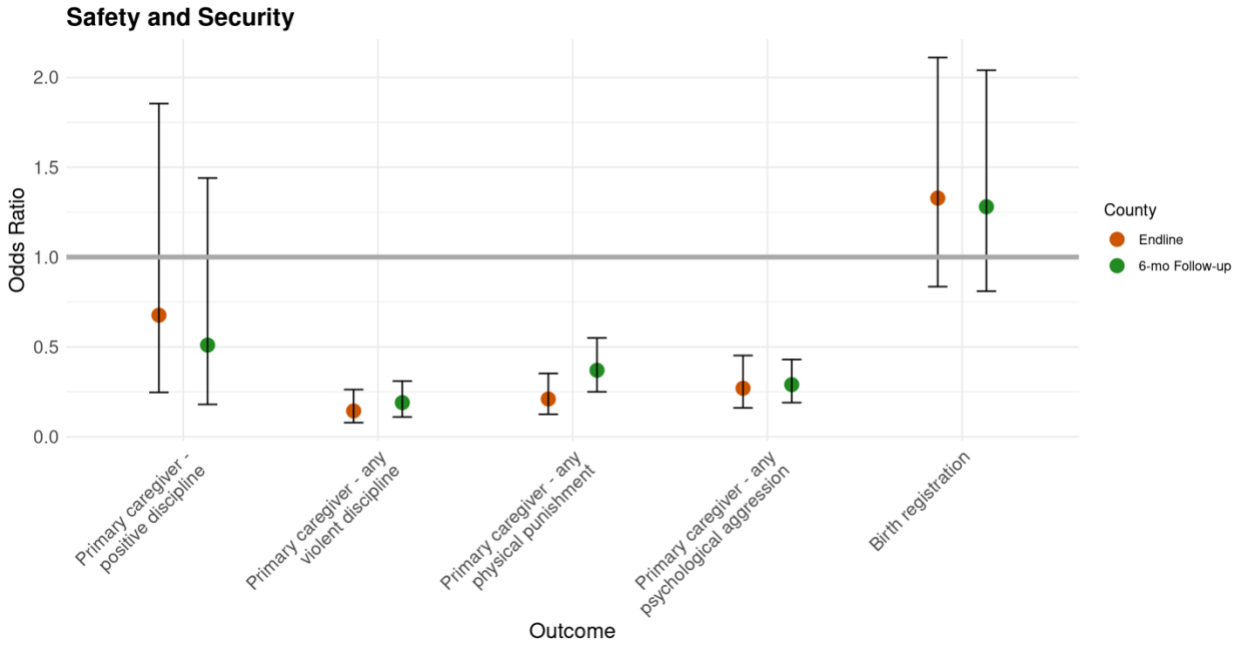


Figure 22. Intervention effects (OR) on safety and security at endline and 6-month follow-up

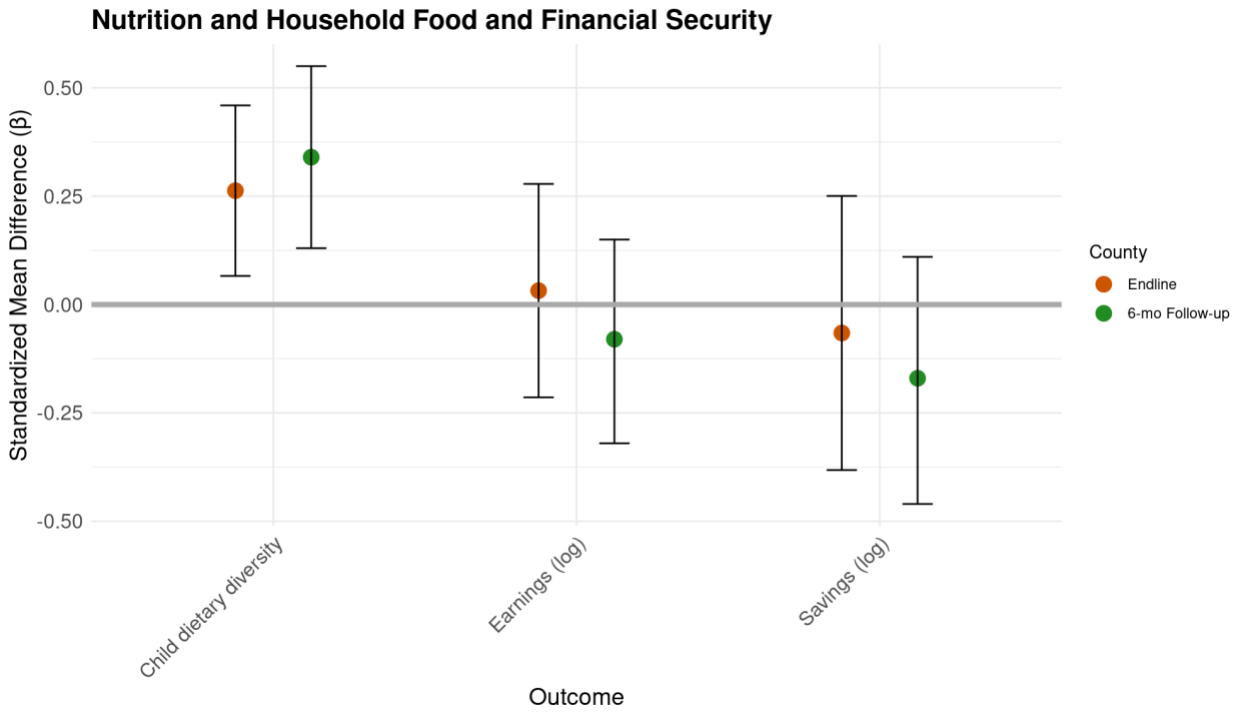


Figure 23. Intervention effects (β) on nutrition, household food, and financial security at endline and 6-month follow-up

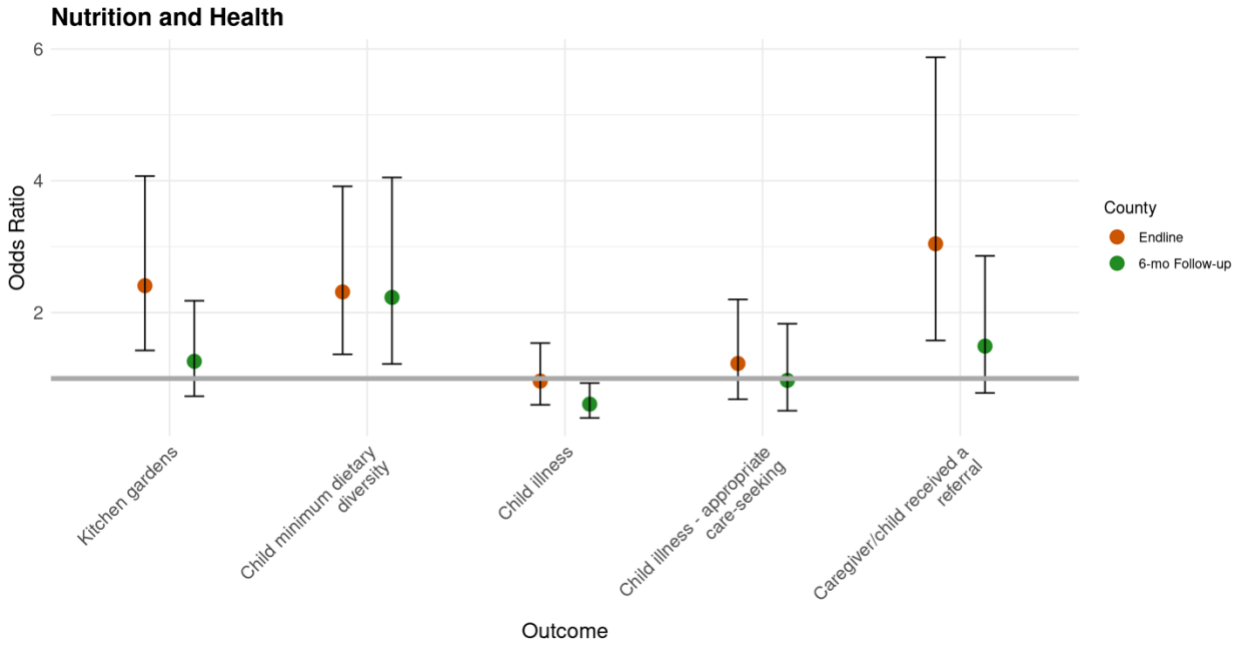


Figure 24. Intervention effects (OR) on nutrition and health at endline and 6-month follow-up

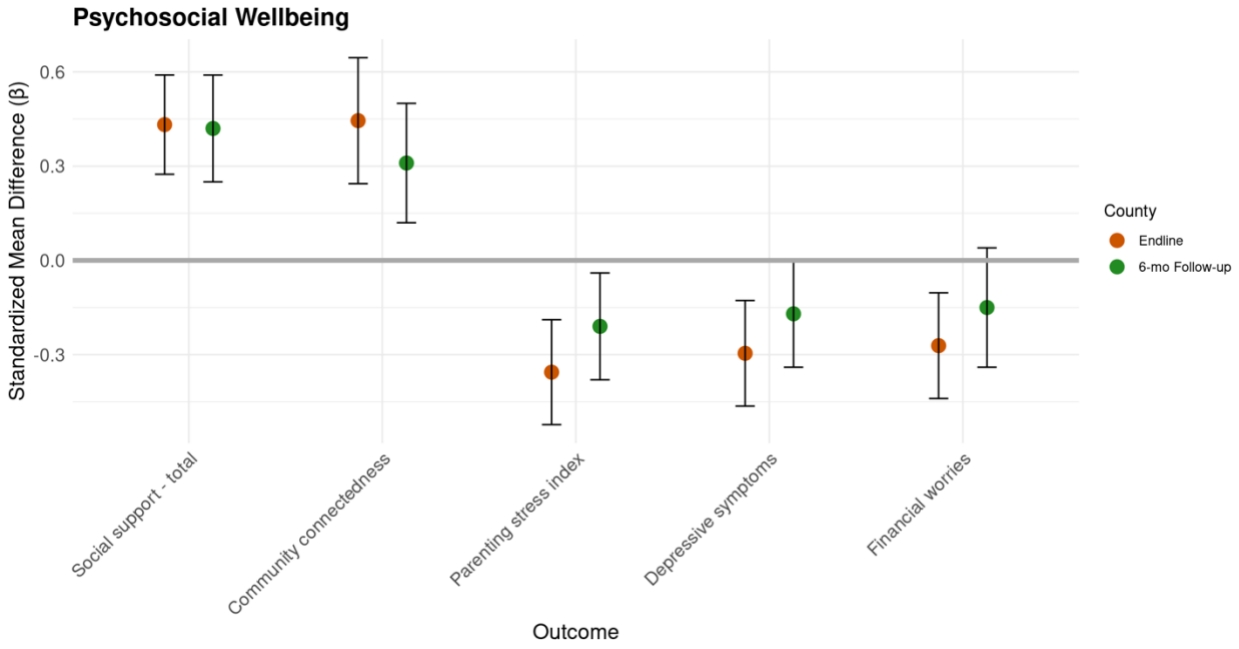


Figure 25. Intervention effects (β) on psychosocial wellbeing at endline and 6-month follow-up

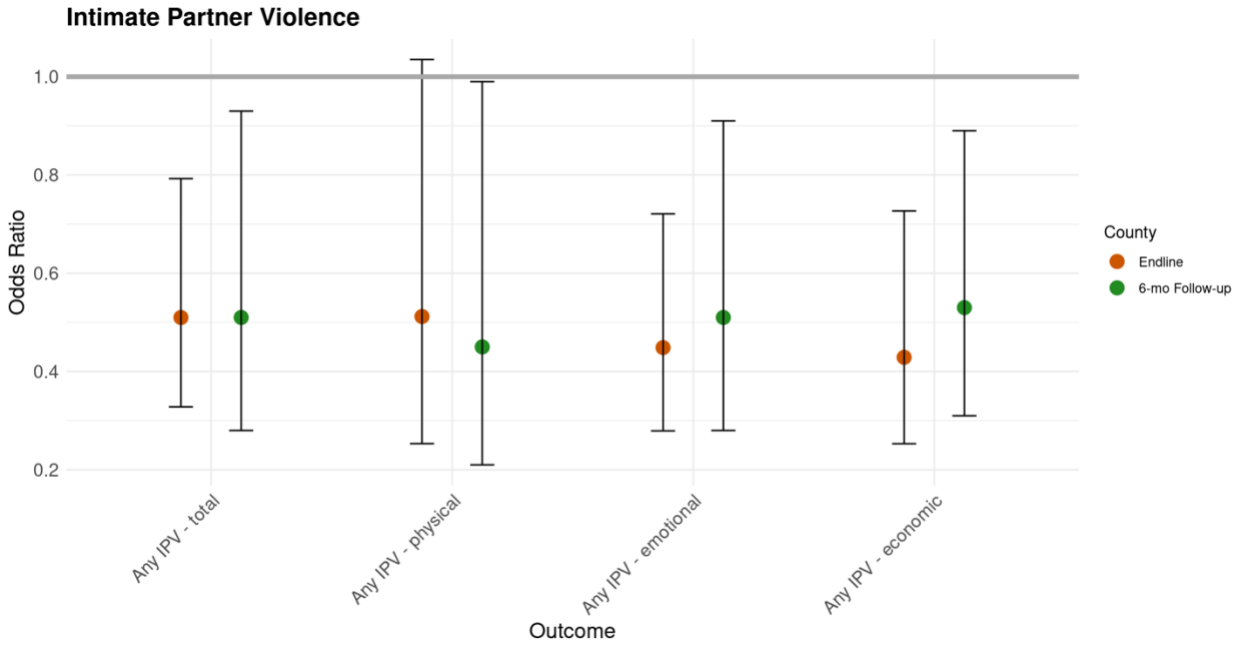


Figure 26. Intervention effects (ORs) on intimate partner violence victimization at endline and 6-month follow-up

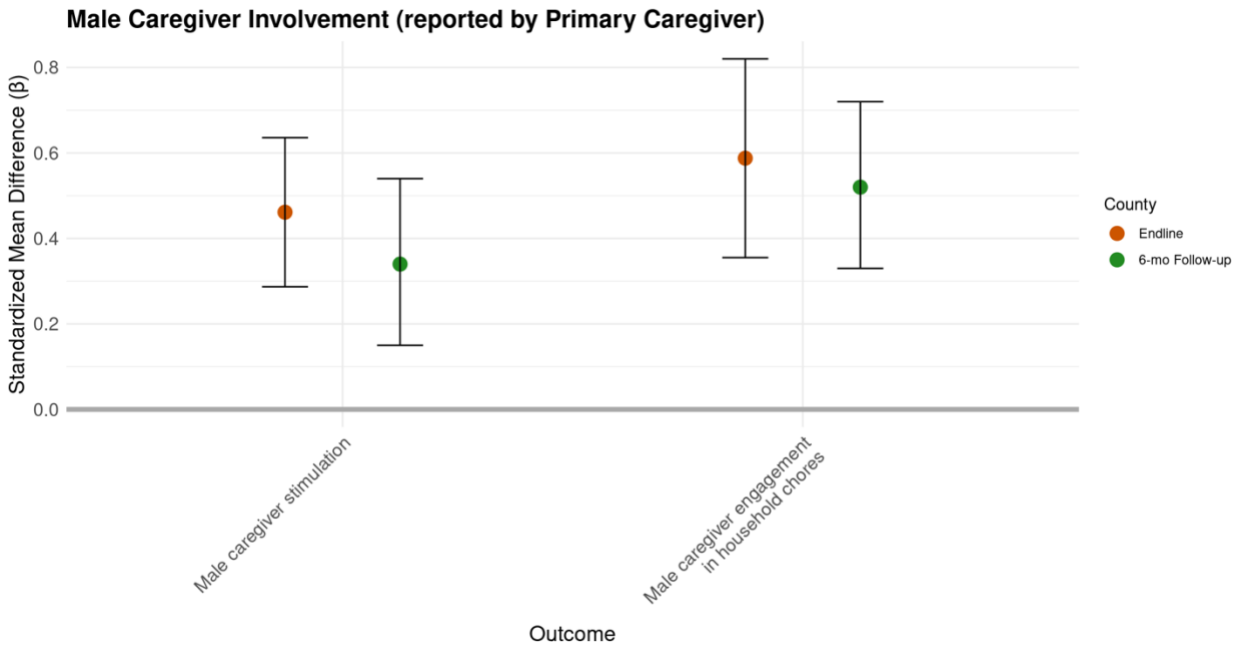


Figure 27. Intervention effects (β) on male caregiver involvement at endline and 6-month follow-up

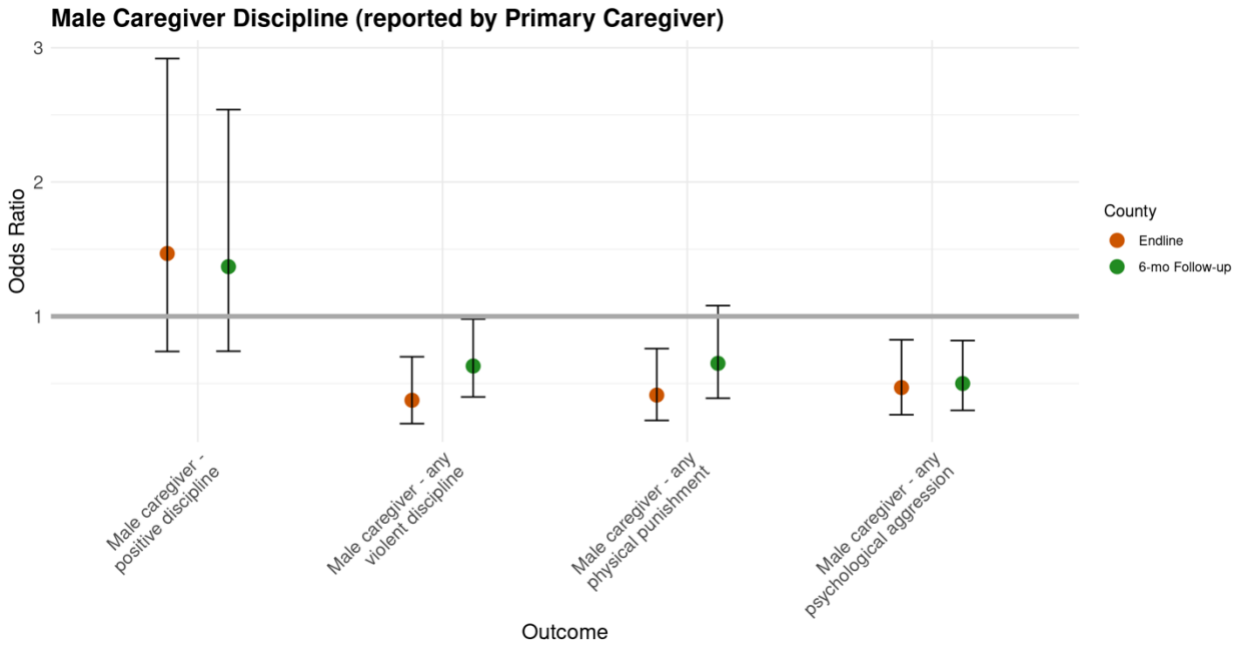


Figure 28. Intervention effects (β) on male caregiver discipline at endline and 6-month follow-up

DISCUSSION

This report presents findings from a six-month follow-up evaluation conducted with the cohort enrolled in the original cluster-randomized controlled trial (RCT) to assess the sustained effects

of the Moments That Matter (MTM) parenting program in Western Kenya. The primary focus of the evaluation was on primary caregiver-level outcomes, including parenting practices, caregiver wellbeing, and other dimensions of nurturing care. While assessing ECD was not the primary aim of the follow-up – given resource constraints and the expectation that many children would have aged out of existing ECD tools – we explored ECD outcomes where feasible. Specifically, we used the CREDI tool to assess developmental outcomes among a subsample of children under 36 months of age (i.e., the maximum age for which the tool is designed and validated). To capture child outcomes across the full age range, we also explored socioemotional and behavioral development using the Strengths and Difficulties Questionnaire (SDQ), which does not have a strict age limitation. Overall, this follow-up assessment provided an opportunity to examine which program benefits were sustained, which may have diminished over time, and reflect on implications for the longer-term potential and durability of MTM's program impact.

At six months post-intervention, several of the caregiver-level outcomes that showed significant improvements at endline were sustained, indicating that many program benefits continued even after the program officially concluded. Primary caregivers in intervention group continued to provide greater early learning opportunities – more stimulation practices, books, and learning materials in the home – compared to those in the control group. Positive parenting also remained significantly better in the intervention group in terms of other caregiving behaviors, including reduced harsh discipline and increased involvement of male caregivers in stimulation and household chores. Beyond parenting practices, mental health benefits for primary caregivers continued to be observed, such as reduced parenting stress and depressive symptoms. While significant reductions in financial worries were noted at endline, these differences were no longer present at the 6-month follow-up. However, the positive effects of MTM in reducing intimate partner violence (IPV) and improving child nutrition, such as dietary diversity, were maintained over time.

What is notable about these results is the breadth of caregiver-level benefits that MTM sustained at least in the short-term across multiple domains. These gains span several thematic pillars of the Nurturing Care Framework, including early learning, safety and security, nutrition, and caregiver mental health. Sustaining improvements in such a diverse range of outcomes six months after program completion speaks to the strength of MTM's holistic design and its ability to support caregivers in the context of raising young children. These multidimensional impacts suggest that a well-integrated approach and through a strong community-based model together with trained ECD promoters and in partnership with faith leaders and local ECD Committees can yield lasting benefits across diverse types of caregiving-related outcomes.

These findings are also consistent with existing literature on parenting interventions, which has shown that program improvements in parenting practices can be maintained over time. Measurement of parental engagement in stimulation is one of the most commonly reassessed outcomes in follow-up studies of parenting programs and has consistently shown stability in effective interventions (Jeong, Pitchik, et al., 2021). Parenting programs have also demonstrated

sustainability in reducing harsh discipline. For example, a meta-analysis found that reductions in family violence can persist up to one or two years post-intervention (Backhaus et al., 2023). On the other hand, sustained mental health effects among caregivers are less commonly observed in the existing literature, even in parenting programs that include explicit psychological components and strategies (e.g., cognitive behavioral therapy, behavioral activation) (Al Sager et al., 2024; Kim et al., 2021). The continued mental health benefits observed in MTM, although modest in size yet statistically significant, suggest a potential area of added value in the program's approach. This finding highlights an important but underexplored domain that warrants further investigation and replication in future studies.

In contrast to the sustained benefits observed for caregivers, MTM had limited if any impacts on ECD outcomes at the six-month follow-up. The only domain that approached statistical significance was socioemotional development, assessed using the CREDI tool for children under 36 months ($\beta = 0.21$ SD, $p = 0.056$). All other CREDI domains showed no statistically significant differences between intervention and control groups at follow-up. Additionally, there were no statistically significant differences in any of the scores based on the SDQ measure (i.e., total difficulties, internalizing problems, externalizing problems, prosocial behaviors), which was administered to the full child sample during this round to capture socioemotional and behavioral outcomes across all ages.

These findings are consistent with the endline evaluation, where ECD effects were modest overall and only observed for child socioemotional scores of the CREDI. Although at endline we had seen significant county-level differences with much larger effect sizes in CREDI scores in Nyamira than Vihiga, by the six-month follow-up, even those earlier gains in Nyamira had largely faded, resulting in no statistically significant differences across counties for any ECD domain. Vihiga county continued to show no effects on ECD outcomes, mirroring the null results at endline. Despite this, county-level differences were observed in a few selected outcomes – specifically in terms of primary caregiver and male caregiver stimulation practices, the availability of books in the home, and referrals – with Nyamira continuing to outperform Vihiga on these indicators at the 6-month follow-up.

These findings add to an evolving evidence base suggesting that while parenting programs can improve caregiver behaviors and even reduce family-level stressors, such changes do not necessarily translate to measurable improvements in child development even when giving it a few months at the completion of the program to take some time to take shape. Several studies, including from Uganda (Atukunda et al., 2019), Pakistan (Yousafzai et al., 2016), and India (Meghir et al., 2023), have found larger and more sustained effects on ECD in short-term follow-ups, and have been the cases where programs achieved medium-to-large immediate gains at endline (>0.4 SD). In contrast, interventions with smaller post-intervention effects on parenting have often failed to generate or sustain impacts on ECD follow-up rounds in the short-run after program completion (Jeong, Pitchik, et al., 2021; Rockers et al., 2018).

The theory of change underpinning most parenting programs posits that strengthening parenting practices, particularly engagement in early learning and responsive caregiving, will lead to measurable improvements in children's developmental outcomes by the end of the program period, which averages around 12 months in program duration in the global evidence. Some proponents suggest that even if ECD effects are not immediately apparent at endline, they may emerge gradually as positive parenting behaviors accumulate and stabilize within the home environment. However, our findings challenge the assumption that any level of improvement is sufficient to generate downstream child impacts. While MTM led to sustained gains in several parenting and caregiver outcomes six months post-intervention, these changes did not translate into meaningful improvements in ECD. This suggests that while caregiving behaviors remain the critical pathway to improving child outcomes, the magnitude, quality, and consistency of these changes may need to be stronger to influence children's developmental trajectories.

Taken together, these findings have important implications for strengthening the MTM program. Enhancing the program's content, delivery approach, and the quality of training and mentorship, particularly in the areas of responsive caregiving and age-appropriate early learning, may help drive larger, more meaningful improvements in parenting behaviors. Such improvements could not only boost caregiver outcomes but also increase the likelihood of developmental gains for children both immediately following the program and in the months that follow, potentially shaping longer-term developmental trajectories. This is especially important given that the magnitude of initial post-intervention effects has been shown to predict sustained child outcomes over time (Jeong, Pitchik, et al., 2021). Additionally, refining how the parenting and ECD content is framed and delivered – by tailoring it to the evolving developmental stages of the child and adapting it over time to match children's changing needs – may further enhance program effectiveness. A curriculum that evolves with the child's ages and stages is more likely to support sustained caregiver engagement and stronger parenting practices to in turn maximize the benefits on ECD outcomes during and beyond the intervention period.

Limitations

It is important to acknowledge several limitations in our study. First, our sample for ECD outcomes using the CREDI tool was limited to children under 36 months of age. By the time of the six-month follow-up, a significant portion of the original sample had aged out of the CREDI, reducing our power to detect differences. Among the younger children who were reassessed with CREDI, only the CREDI-socioemotional development scores showed an effect that approached statistical significance. Moreover, many of our outcome measures were based on caregiver self-report, raising the potential for recall or social desirability bias. While we used validated tools, incorporating direct observations or performance-based assessments of ECD in future evaluations would enhance the objectivity of findings. Lastly, approximately 15% of children

from the original cohort were lost to follow-up. However, there were no meaningful baseline differences between those lost and retained, minimizing the risk of attrition bias.

CONCLUSION

Six months after the end of the MTM parenting program, many positive caregiver-level outcomes were maintained (e.g., increased stimulation, reduced harsh discipline, improved mental health). These sustained changes are encouraging and affirm the value of investing in parenting programs that prioritize the holistic wellbeing of caregivers and families. However, the limited and diminishing effects on early child development, even in the context of improved parenting behaviors, signal a need for program strengthening.

As interest in scaling parenting interventions grows globally, evidence of sustained benefits especially in the immediate months after program completion and specifically on child-level outcomes, given that young children are the intended ultimate beneficiary or target person in the theory of change of parenting programs, is increasingly seen as a powerful signal of real and enduring impact. For MTM, this represents both a challenge and an opportunity. Rethinking and refining the program model to better support lasting improvements in parenting and child development is not only feasible, but also a worthwhile and strategic goal. The absence of meaningful ECD effects at the 6-month follow-up suggests that targeted improvements to the program's content, delivery strategy, or overall design – such as embedding stronger and more sustained support for early learning and responsive caregiving – could help MTM better realize its full potential. Strengthening these components would not only benefit caregivers but also increase the likelihood of achieving lasting, positive impacts on children's developmental outcomes, which is a core marker of program success and a critical foundation for supporting children to thrive across their life course.

REFERENCES

Al Sager, A., Goodman, S. H., Jeong, J., Bain, P. A., & Ahun, M. N. (2024). Effects of multi-component parenting and parental mental health interventions on early childhood development and parent outcomes: a systematic review and meta-analysis. *Lancet Child Adolesc Health*, *8*(9), 656-669. https://doi.org/10.1016/S2352-4642(24)00134-2

Atukunda, P., Muhoozi, G. K. M., van den Broek, T. J., Kort, R., Diep, L. M., Kaaya, A. N., Iversen, P. O., & Westerberg, A. C. (2019). Child development, growth and microbiota: follow-up of a randomized education trial in Uganda. *J Glob Health*, *9*(1), 010431. https://doi.org/10.7189/jogh.09.010431

Backhaus, S., Leijten, P., Jochim, J., Melendez-Torres, G. J., & Gardner, F. (2023). Effects over time of parenting interventions to reduce physical and emotional violence against children: a systematic review and meta-analysis. *EClinicalMedicine*, *60*, 102003. https://doi.org/10.1016/j.eclinm.2023.102003

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., Perez-Escamilla, R., Rao, N., Ip, P., Fernald, L. C. H., MacMillan, H., Hanson, M., Wachs, T. D., Yao, H., Yoshikawa, H., Cerezo, A., Leckman, J. F., Bhutta, Z. A., & Early Childhood Development Interventions Review Group, f. t. L. E. C. D. S. S. C. (2017). Nurturing care: promoting early childhood development. *Lancet*, *389*(10064), 91-102. https://doi.org/10.1016/S0140-6736(16)31390-3

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(9), 1159-1167. https://doi.org/10.1097/01.chi.0000132808.36708.a9

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S. M., & Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, *344*(6187), 998-1001. https://doi.org/doi:10.1126/science.1251178

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to Use Broader Internalising and Externalising Subscales Instead of the Hypothesised Five Subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British Parents, Teachers and Children. *Journal of Abnormal Child Psychology*, *38*(8), 1179-1191. https://doi.org/10.1007/s10802-010-9434-x

Jeong, J., Franchett, E. E., Ramos de Oliveira, C. V., Rehmani, K., & Yousafzai, A. K. (2021). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS Med*, *18*(5), e1003602. https://doi.org/10.1371/journal.pmed.1003602

Jeong, J., Pitchik, H. O., & Fink, G. (2021). Short-term, medium-term and long-term effects of early parenting interventions in low- and middle-income countries: a systematic review. *BMJ Glob Health*, *6*(3). https://doi.org/10.1136/bmjgh-2020-004067

Justino, P., Leone, M., Rolla, P., Abimpaye, M., Dusabe, C., Uwamahoro, M. D., & Germond, R. (2023). Improving Parenting Practices for Early Child Development: Experimental Evidence from Rwanda. *Journal of the European Economic Association*, *21*(4), 1510-1550. https://doi.org/10.1093/jeea/jvac070

Kim, E. T., Opiyo, T., Acayo, P. S., Lillie, M., Gallis, J., Zhou, Y., Ochieng, M., Okuro, S., Hembling, J., McEwan, E., & Baumgartner, J. N. (2021). Effect of a lay counselor delivered integrated maternal mental health and early childhood development group-based intervention in Siaya County, Kenya: A quasi-experimental longitudinal study. *Journal of Affective Disorders*, *292*, 284-294. https://doi.org/https://doi.org/10.1016/j.jad.2021.06.002

Meghir, C., Attanasio, O., Jervis, P., Day, M., Makkar, P., Behrman, J., Gupta, P., Pal, R., Phimister, A., Vernekar, N., & Grantham-McGregor, S. (2023). Early Stimulation and Enhanced Preschool: A Randomized Trial. *Pediatrics*, *151*(Supplement 2), e2023060221H. https://doi.org/10.1542/peds.2023-060221H

Rockers, P. C., Zanolini, A., Banda, B., Chipili, M. M., Hughes, R. C., Hamer, D. H., & Fink, G. (2018). Two-year impact of community-based health screening and parenting groups on child development in Zambia: Follow-up to a cluster-randomized controlled trial. *PLOS Medicine*, *15*(4), e1002555. https://doi.org/10.1371/journal.pmed.1002555

Walker, S. P., Chang, S. M., Wright, A. S., Pinto, R., Heckman, J. J., & Grantham-McGregor, S. M. (2022). Cognitive, psychosocial, and behaviour gains at age 31 years from the Jamaica early childhood stimulation trial. *Journal of Child Psychology and Psychiatry*, *63*(6), 626-635. https://doi.org/https://doi.org/10.1111/jcpp.13499

Yousafzai, A. K., Obradovic, J., Rasheed, M. A., Rizvi, A., Portilla, X. A., Tirado-Strayer, N., Siyal, S., & Memon, U. (2016). Effects of responsive stimulation and nutrition interventions on children's development and growth at age 4 years in a disadvantaged population in Pakistan: a longitudinal follow-up of a cluster-randomised factorial effectiveness trial. *Lancet Glob Health*, *4*(8), e548-558. https://doi.org/10.1016/S2214-109X(16)30100-0

APPENDICES

Appendix 1. Descriptive statistics of outcomes, Nyamira county.

	Baseline			Endline			6-Month Follow-up		
	Control n=143	Intervention n=154	p-value	Control n=128	Intervention n=136	p-value	Control n=116	Intervention n=128	p-value
CREDI Overall score	44.8 (3.9)	44.9 (3.7)	0.969	51.7 (1.5)	52.3 (1.4)	**0.006**	52.5 (1.1)	52.6 (0.9)	0.451
CREDI Cognition score	47.4 (2.3)	47.3 (2.2)	0.936	50.7 (1.0)	51.2 (0.9)	**0.002**	51.2 (0.9)	51.3 (0.7)	0.486
CREDI Language score	47.6 (1.7)	47.6 (1.7)	0.996	51.4 (1.2)	51.9 (1.2)	**0.008**	52.1 (1.0)	52.2 (0.9)	0.509
CREDI Motor score	46.6 (2.5)	46.7 (2.5)	0.869	51.6 (1.3)	51.8 (1.2)	**0.126**	52.0 (1.0)	52.0 (0.8)	0.801
CREDI Social-emotional score	47.1 (2.5)	47.0 (2.4)	0.708	51.3 (1.0)	51.9 (1.0)	**<0.001**	52.0 (0.9)	52.1 (0.7)	0.208
Primary caregiver stimulation index score (11 item)	6.8 (2.5)	7.1 (2.6)	0.371	8.9 (2.2)	9.9 (2.2)	**<0.001**	7.4 (3.4)	9.6 (2.4)	**<0.001**
Number of learning materials	2.3 (1.7)	2.6 (1.8)	0.192	4.0 (1.1)	4.5 (0.8)	**<0.001**	3.6 (1.1)	4.2 (0.9)	**<0.001**
Number of books in household	0.7 (1.7)	0.8 (1.5)	0.751	1.8 (1.9)	2.7 (2.1)	**<0.001**	1.3 (1.7)	2.4 (2.9)	**<0.001**
Primary caregiver use of any positive discipline	97 (68%)	101 (66%)	0.681	121 (97%)	124 (95%)	0.561	108 (93 %)	122 (95%)	0.459
Primary caregiver use of any violent discipline	86 (60%)	87 (56%)	0.524	107 (86%)	54 (42%)	**<0.001**	101 (87%)	79 (62%)	**<0.001**
Primary caregiver use of any physical punishment	76 (53%)	73 (47%)	0.323	97 (78%)	48 (37%)	**<0.001**	85 (73%)	66 (52%)	**<0.001**
Primary caregiver use of any psychological aggression	73 (51%)	70 (45%)	0.335	83 (66%)	36 (28%)	**<0.001**	89 (77%)	61 (48%)	**<0.001**
Birth registration	48 (34%)	71 (46%)	0.031	69 (56%)	99 (77%)	**<0.001**	61 (53%)	98 (77%)	**<0.001**
Kitchen garden in household	112 (78%)	110 (71%)	0.172	99 (79%)	109 (84%)	0.339	86 (74%)	101 (79%)	0.379
Child dietary diversity score (24 hr)	3.9 (1.6)	3.6 (1.5)	0.182	4.2 (1.3)	4.8 (1.2)	**0.001**	4.2 (1.4)	4.9 (1.4)	**<0.001**
Children aged 6+ months who receive a minimum dietary diversity (MDD) in past 24 hours	33 (36%)	30 (32%)	0.604	49 (41%)	71 (58%)	**0.007**	44 (42%)	82 (66%)	**<0.001**
Child experienced any illness (diarrhea, cough, or fever) in last 2 weeks	101 (71%)	118 (77%)	0.241	54 (44%)	70 (55%)	0.088	84 (74%)	80 (63%)	0.076

Appropriate care (hospital, clinic, CHV) sought for any child illness in past 2 weeks	45 (45%)	80 (68%)	**<0.001**	22 (41%)	45 (64%)	**0.009**	39 (46%)	56 (70%)	**0.002**
Caregiver/child received a referral in past 6 months				12 (10%)	48 (37%)	**0.000**	12 (10%)	31 (24%)	**0.005**
Income in past month (KSH)	1431.5 (3271.7)	1721.4 (4093.9)	0.503	3444.9 (5568.8)	2741.0 (4765.5)	0.332	5001.7 (6824.5)	4192.1 (7990.2)	0.534
Total amount currently in savings (i.e Bank, SACCO etc) (KSH)	3642.0 (25404.5)	982.5 (2829.4)	0.198	3456.0 (9792.6)	3757.3 (6292.4)	0.769	4485.4 (14671.9)	3553.9 (5096.4)	0.500
Money accessed in credit in past month (KSH)	27229.0 (129972.1)	45712.6 (257903.6)	0.614	15074.7 (34782.7)	7864.3 (16172.3)	0.185	19534.4 (34716.7)	7030.8 (10891.9)	**0.010**
Primary caregiver overall social support total score	3.4 (0.7)	3.4 (0.6)	0.939	3.4 (0.8)	3.6 (0.6)	**0.011**	3.4 (0.8)	3.6 (0.6)	**0.045**
Community connectedness mean score	3.2 (0.9)	3.3 (0.9)	0.662	3.0 (1.3)	3.4 (1.2)	**0.001**	3.1 (1.0)	3.5 (0.9)	**0.002**
Primary caregiver - any IPV victimization (physical, emotional or economic)	56 (47%)	66 (49%)	0.776	37 (38%)	33 (29%)	0.145	35 (38%)	25 (32%)	0.331
Primary caregiver - any physical IPV victimization	18 (15%)	30 (22%)	0.150	13 (13%)	10 (9%)	0.272	8 (9%)	6 (5%)	0.357
Primary caregiver - any emotional IPV victimization	47 (40%)	58 (43%)	0.579	30 (31%)	28 (24%)	0.284	32 (35%)	28 (25%)	0.137
Primary caregiver - any economic IPV victimization	40 (34%)	38 (28%)	0.342	22 (23%)	17 (15%)	0.139	20 (22%)	20 (18%)	0.507
Primary caregiver parenting stress total score	32.6 (7.1)	32.3 (7.3)	0.742	33.3 (6.8)	30.9 (6.5)	**0.004**	35.4 (5.0)	34.4 (5.4)	0.129
Primary caregiver depression total score	7.3 (5.8)	8.0 (6.5)	0.283	8.9 (5.9)	7.7 (5.6)	0.100	8.5 (6.4)	8.6 (6.5)	0.880
Primary caregiver - financial worries in past month				12.8 (5.0)	11.6 (5.4)	0.059	10.7 (5.5)	10.4 (5.5)	0.611
Father stimulation index score (11 item)	5.0 (3.3)	5.2 (3.5)	0.581	6.4 (4.2)	8.3 (3.9)	**<0.001**	4.8 (3.9)	6.8 (4.5)	**0.001**
Father use of any positive discipline	45 (42%)	50 (43%)	0.828	94 (86%)	95 (85%)	0.765	60 (64%)	84 (77%)	**0.038**
Father use of any violent discipline	34 (31%)	37 (32%)	0.947	60 (55%)	23 (21%)	**<0.001**	37 (39%)	34 (31%)	0.224
Father use of any psychological aggression	25 (23%)	22 (19%)	0.442	37 (34%)	16 (14%)	**<0.001**	28 (30%)	25 (23%)	0.268
Father use of any physical punishment	28 (26%)	32 (28%)	0.779	49 (45%)	16 (14%)	**<0.001**	32 (34%)	26 (24%)	0.109

Father involvement in household chores subscale score	2.3 (2.4)	2.3 (2.5)	0.953	1.3 (1.9)	3.0 (2.6)	**<0.001**	1.3 (1.9)	2.9 (2.6)	**<0.001**

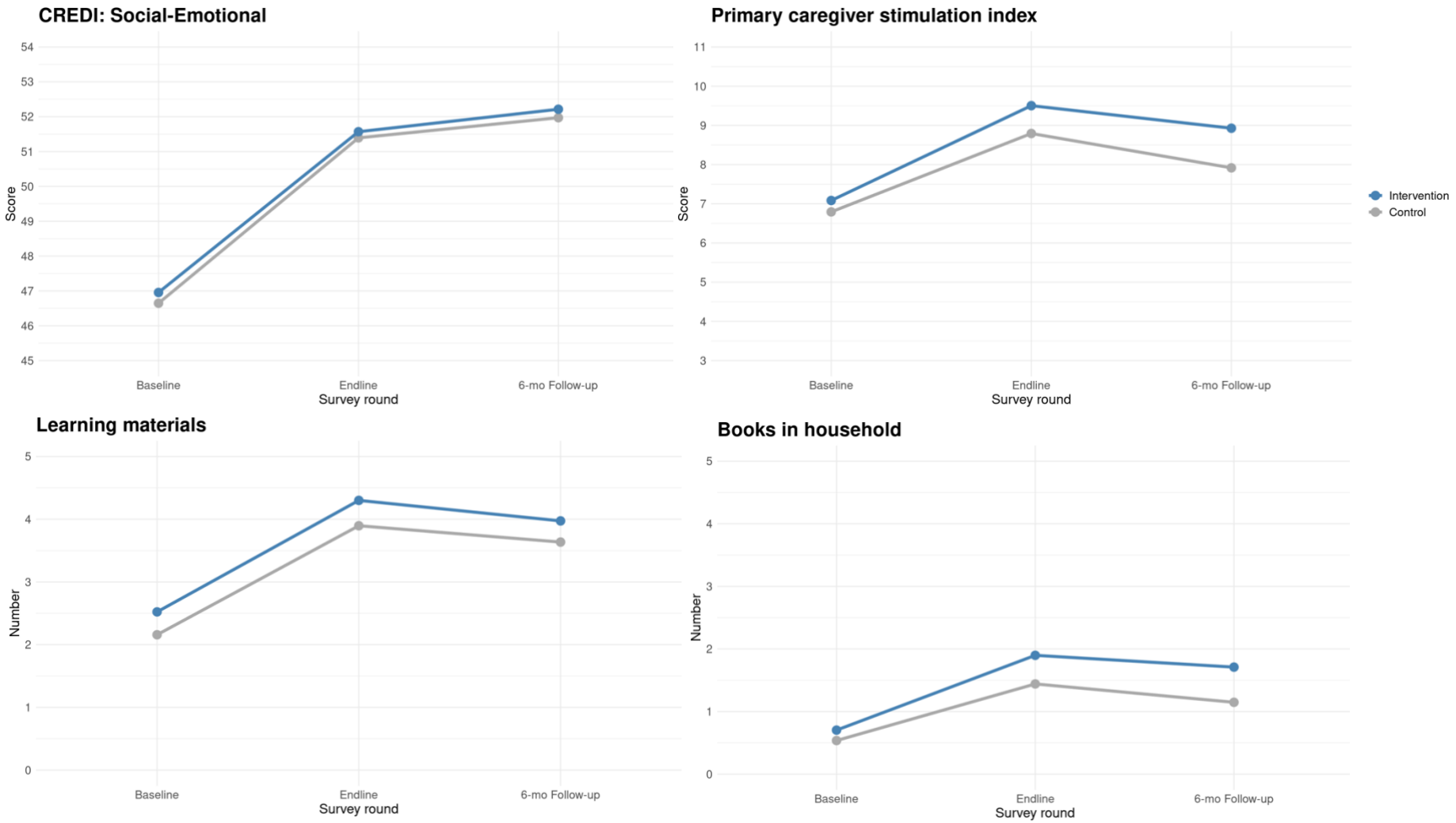
Appendix 2. Descriptive statistics of outcomes, Vihiga county.

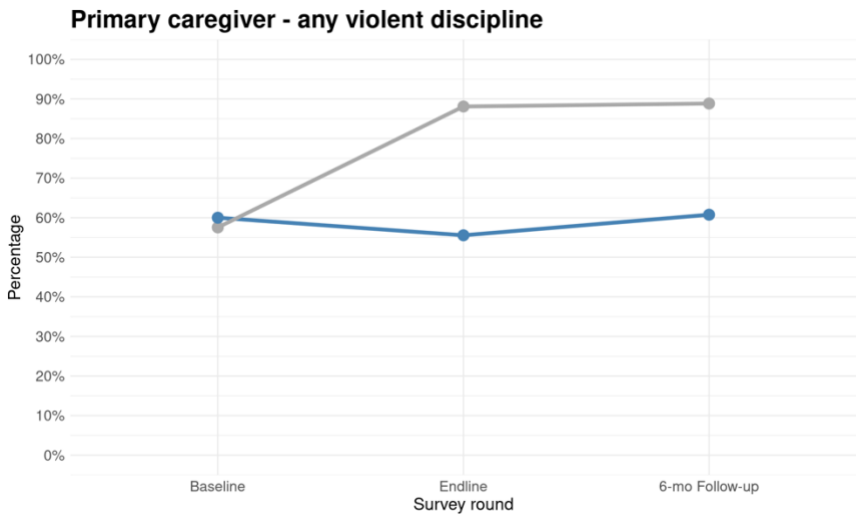
	Baseline			**Endline**			**6-Month Follow-up**		
	Control n=142	Intervention n=156	P>|z|	Control n=128	Intervention n=133	P>|z|	Control n=119	Intervention n=123	p-value
CREDI Overall score	43.6 (3.9)	44.6 (4.0)	**0.025**	51.9 (1.5)	52.0 (1.6)	0.855	52.5 (1.1)	52.7 (1.3)	0.466
CREDI Cognition score	46.5 (2.3)	47.2 (2.2)	**0.019**	50.8 (1.0)	50.7 (1.0)	0.814	51.1 (0.8)	51.2 (0.9)	0.547
CREDI Language score	47.0 (1.6)	47.4 (1.7)	**0.037**	51.8 (1.2)	51.8 (1.2)	0.820	52.1 (1.0)	52.2 (1.1)	0.446
CREDI Motor score	46.0 (2.6)	46.6 (2.6)	**0.044**	51.4 (1.2)	51.5 (1.3)	0.519	51.9 (0.9)	51.9 (1.2)	0.913
CREDI Social-emotional score	46.2 (2.5)	46.9 (2.5)	**0.012**	51.4 (1.2)	51.4 (1.1)	0.982	52.0 (0.8)	52.2 (0.9)	0.126
Primary caregiver stimulation index score (11 item)	6.7 (2.7)	7.1 (2.5)	0.291	8.7 (2.0)	9.2 (2.6)	0.120	8.4 (2.7)	8.5 (3.1)	0.779
Number of learning materials	2.0 (1.7)	2.4 (1.7)	0.017	3.8 (1.0)	4.1 (1.0)	0.028	3.6 (1.1)	3.8 (1.3)	0.432
Number of books in household	0.4 (1.2)	0.6 (1.5)	0.079	1.1 (1.5)	1.3 (1.6)	0.292	1.0 (1.5)	1.0 (1.1)	0.841
Primary caregiver use of any positive discipline	55 (39%)	71 (46%)	0.237	124 (98%)	123 (94%)	0.137	119 (100%)	114 (93%)	**0.003**
Primary caregiver use of any violent discipline	78 (55%)	99 (63%)	0.134	115 (91%)	90 (69%)	**<0.001**	107 (89%)	7 (63%)	**<0.001**
Primary caregiver use of any physical punishment	64 (45%)	83 (53%)	0.161	103 (81%)	78 (60%)	**<0.001**	92 (77%)	69 (56%)	**<0.001**
Primary caregiver use of any psychological aggression	57 (40%)	72 (46%)	0.295	90 (71%)	66 (50%)	**<0.001**	83 (70%)	57 (46%)	**<0.001**
Birth registration	39 (28%)	38 (24%)	0.517	75 (59%)	65 (50%)	0.128	55 (46%)	35 (28%)	**0.004**
Kitchen garden in household	94 (66%)	100 (64%)	0.705	85 (67%)	112 (85%)	**<0.001**	82 (69%)	80 (65%)	0.523
Child dietary diversity score (24 hr)	3.1 (1.5)	3.1 (1.6)	0.811	3.6 (1.4)	3.9 (1.4)	0.080	3.2 (1.5)	3.6 (1.6)	0.063
Children aged 6+ months who receive a minimum dietary diversity (MDD) in past 24 hours	13 (18%)	25 (27%)	0.211	32 (26%)	42 (33%)	0.130	25 (21%)	33 (28%)	0.198

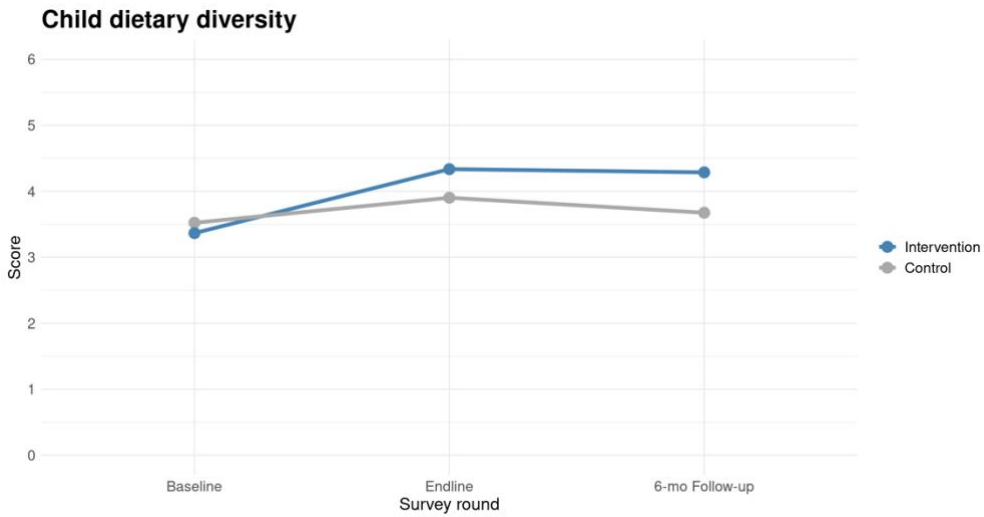
Child experienced any illness (diarrhea, cough, or fever) in last 2 weeks	105 (74%)	123 (79%)	0.319	89 (70%)	73 (58%)	**0.044**	94 (79%)	88 (73%)	0.257
Appropriate care (hospital, clinic, CHV) sought for any child illness in past 2 weeks	55 (52%)	70 (57%)	0.493	42 (47%)	32 (44%)	0.670	32 (34%)	25 (28%)	0.413
Caregiver/child received a referral in past 6 months				23 (18%)	30 (23%)	0.341	20 (17%)	12 (10%)	0.106
Income in past month (KSH)	680.6 (1332.0)	862.2 (3302.9)	0.541	2936.8 (4664.0)	3527.5 (5088.6)	0.473	3241.8 (2363.8)	3407.5 (3593.0)	0.765
Total amount currently in savings (i.e Bank, SACCO etc) (KSH)	131.3 (518.7)	558.5 (3733.6)	0.178	968.5 (3381.7)	1912.8 (3182.2)	**0.022**	1187.0 (5383.1)	1592.0 (3357.8)	0.482
Money accessed in credit in past month (KSH)	2503.6 (3704.2)	4289.2 (15100.4)	0.602	5365.5 (8945.1)	2875.8 (2118.2)	**0.037**	10795.5 (23074.7)	3248.9 (3647.4)	**0.031**
Primary caregiver overall social support total score	3.3 (0.6)	3.3 (0.6)	0.744	3.2 (0.8)	3.5 (0.8)	**0.001**	3.3 (0.9)	3.6 (0.8)	**0.003**
Community connectedness mean score	2.8 (1.0)	2.8 (0.9)	0.414	2.9 (0.9)	3.3 (0.9)	**0.001**	2.8 (1.3)	3.0 (1.3)	**0.144**
Primary caregiver - any IPV victimization (physical, emotional or economic)	54 (46%)	54 (46%)	0.952	57 (55%)	34 (34%)	**0.003**	46 (47%)	38 (40%)	0.272
Primary caregiver - any physical IPV victimization	19 (16%)	17 (14%)	0.697	17 (16%)	10 (10%)	0.181	17 (18%)	11 (11%)	0.231
Primary caregiver - any emotional IPV victimization	48 (41%)	47 (40%)	0.852	48 (46%)	24 (24%)	**<0.001**	38 (39%)	33 (34%)	0.489
Primary caregiver - any economic IPV victimization	36 (31%)	33 (28%)	0.637	37 (36%)	17 (17%)	**0.003**	33 (34%)	17 (17%)	**0.010**
Primary caregiver parenting stress total score	33.3 (5.7)	35.5 (6.7)	**0.003**	33.4 (8.2)	31.0 (29.6)	**0.019**	36.1 (5.7)	35.3 (6.3)	0.295
Primary caregiver depression total score	9.2 (5.8)	10.4 (6.7)	0.089	9.8 (8.3)	7.9 (6.9)	**0.048**	9.6 (7.4)	8.8 (6.7)	0.361
Primary caregiver - financial worries in past month				12.9 (5.1)	11.8 (5.4)	0.088	13.8 (4.4)	13.8 (4.0)	0.989
Father stimulation index score (11 item)	3.5 (3.3)	4.3 (3.5)	0.059	5.4 (3.7)	6.8 (4.0)	**0.014**	5.3 (4.0)	6.0 (4.1)	0.230
Father use of any positive discipline	22 (20%)	32 (29%)	0.118	82 (74%)	79 (76%)	0.724	82 (80%)	75 (78%)	**0.797**
Father use of any violent discipline	32 (29%)	35 (31%)	0.662	49 (44%)	35 (34%)	0.115	48 (47%)	36 (38%)	0.194
Father use of any psychological aggression	20 (18%)	20 (18%)	1.000	36 (32%)	23 (22%)	0.090	32 (31%)	19 (20%)	0.069

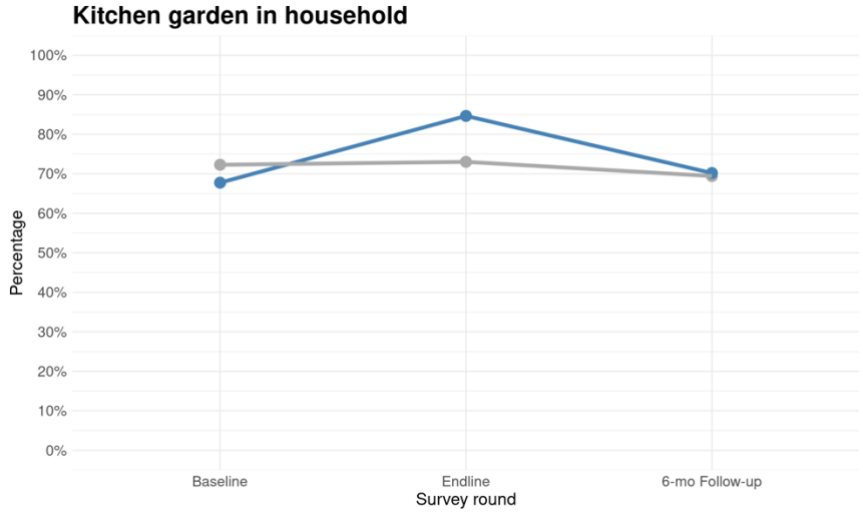
Father use of any physical punishment	23 (21%)	29 (26%)	0.342	32 (29%)	25 (24%)	0.426	35 (34%)	28 (29%)	0.466
Father involvement in household chores subscale score	2.1 (2.2)	2.0 (2.2)	0.676	0.9 (1.6)	1.3 (1.9)	0.188	1.4 (1.8)	1.8 (2.1)	0.123

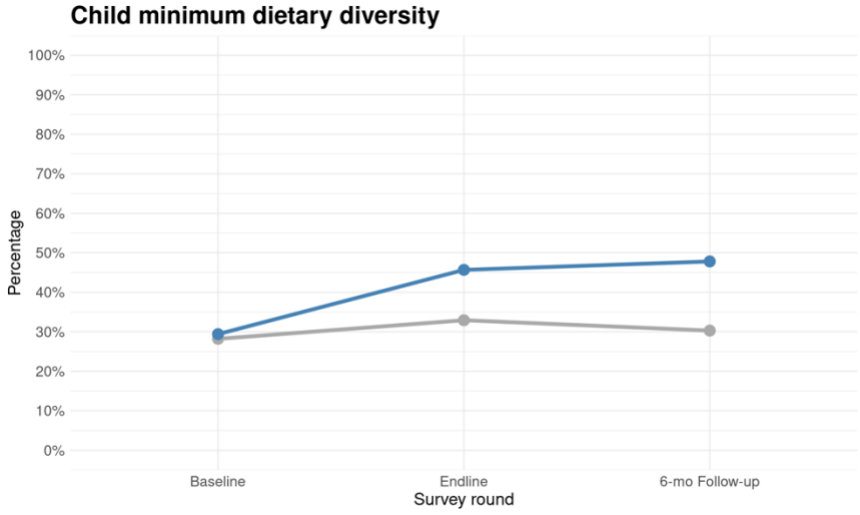
Appendix 3. Descriptive statistics (means of %) of outcomes for intervention and control groups over time (at baseline, endline, and 6-month follow-up).

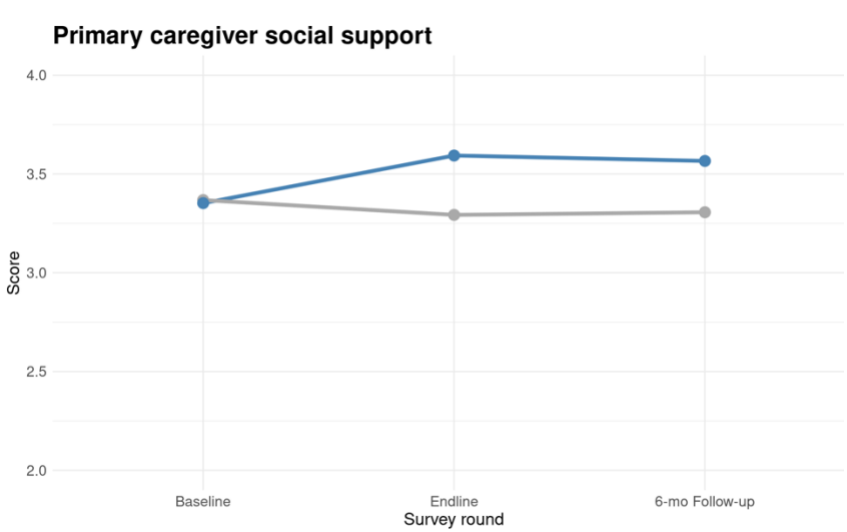


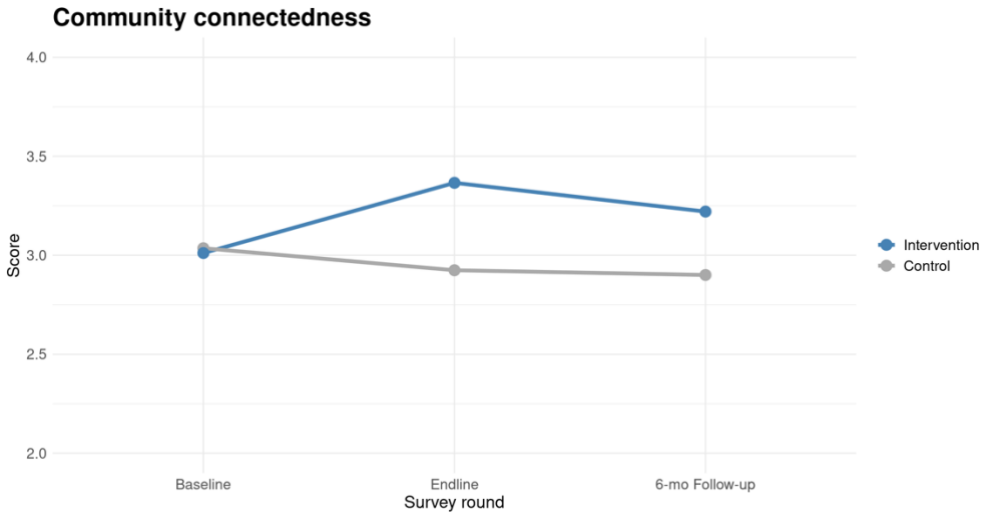

Primary caregiver - any violent discipline
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
Baseline
Endline
6-mo Follow-up
Survey round

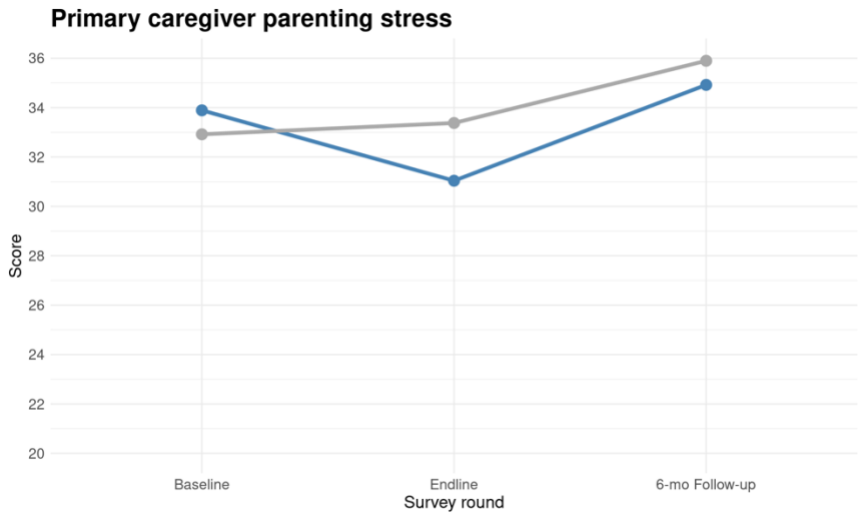

Child dietary diversity
6
5
4
3
2
1
0
Score
Baseline
Endline
6-mo Follow-up
Survey round
Intervention
Control

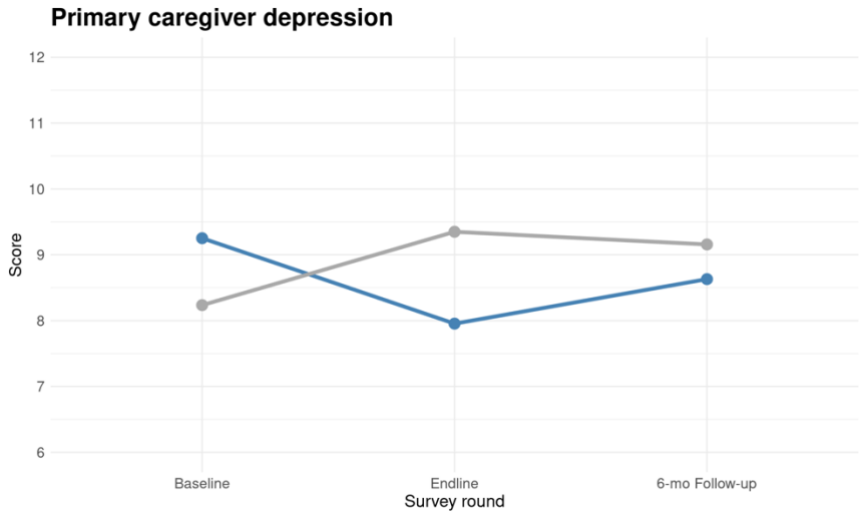

Kitchen garden in household
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
Baseline
Endline
6-mo Follow-up
Survey round

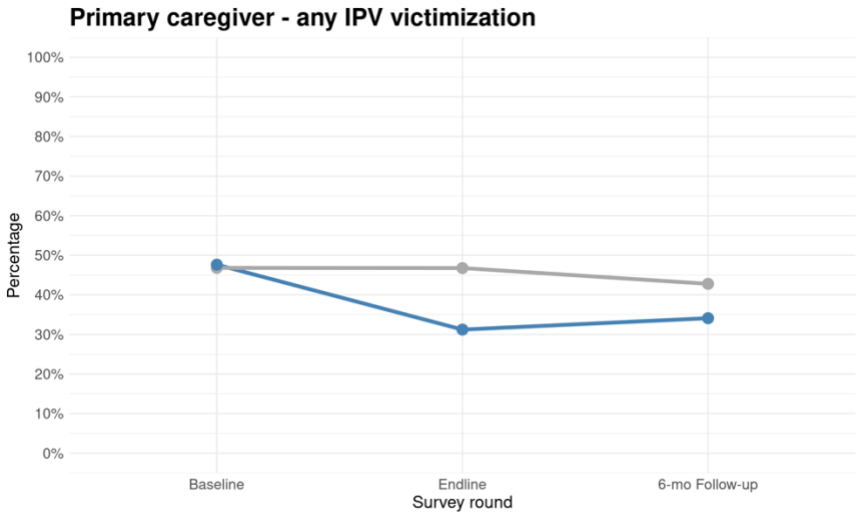

Child minimum dietary diversity
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
Baseline
Endline
6-mo Follow-up
Survey round

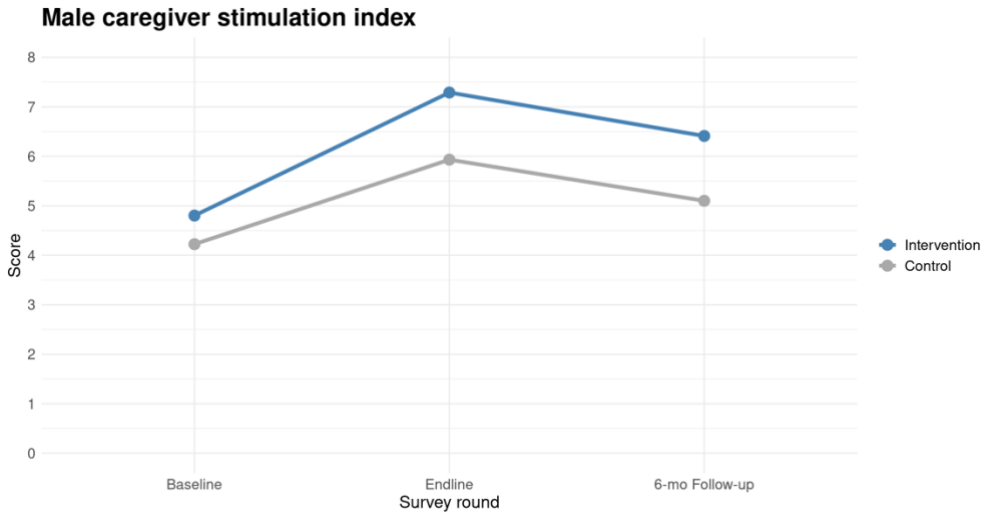

Primary caregiver social support
4.0
3.5
3.0
2.5
2.0
Score
Baseline
Endline
6-mo Follow-up
Survey round

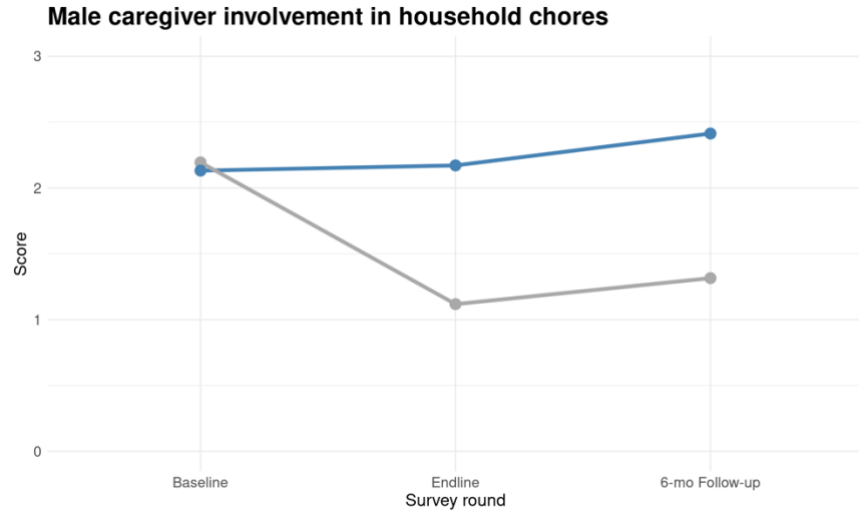

Community connectedness
4.0
3.5
3.0
2.5
2.0
Score
Baseline
Endline
6-mo Follow-up
Survey round
Intervention
Control

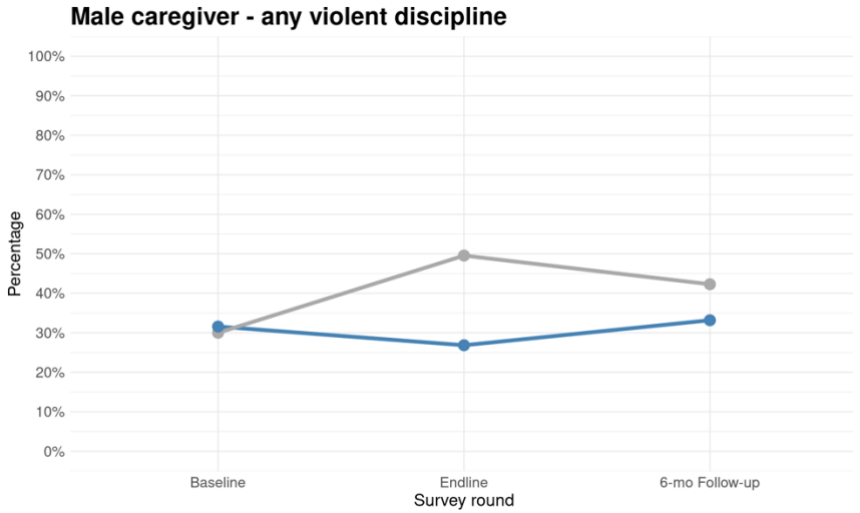

Primary caregiver parenting stress
36
34
32
30
28
26
24
22
20
Score
Baseline
Endline
6-mo Follow-up
Survey round


Primary caregiver depression
12
11
10
9
8
7
6
Score
Baseline
Endline
6-mo Follow-up
Survey round


Primary caregiver - any IPV victimization
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
Baseline
Endline
6-mo Follow-up
Survey round


Male caregiver stimulation index
8
7
6
5
4
3
2
1
0
Score
Baseline
Endline
6-mo Follow-up
Survey round
Intervention
Control


Male caregiver involvement in household chores
3
2
1
0
Score
Baseline
Endline
6-mo Follow-up
Survey round


Male caregiver - any violent discipline
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
Baseline
Endline
6-mo Follow-up
Survey round