Effectiveness Trial of a Family Home Visiting Intervention to Promote Early Child Development Among Families Served by the Social Protection System in Rwanda

NCT02510313

Finalized: October 8, 2019

A. Background

Introduction

Rwanda is invested in addressing indicators of underdevelopment in children. Rwanda recognizes that affordable, scalable, and effective interventions are needed to advance the prospects of extremely vulnerable children (Engle et al., 2011, Britto et al., 2017). In particular, children living in poverty are at risk for poor health and development in addition to exposure to family stress and violence, which can perpetuate intergenerational cycles of violence, poverty, and lost human capital (Richter et al., 2017; Shonkoff & Fisher, 2013). Through its flagship social protection program Vision 2020 Umurenge Program (VUP), which targets the most vulnerable families in Rwanda (those with an Ubudehe 1 or 2 classification) the Rwandan government is working to address family violence, promotion of gender equality and increased attention to social safety nets. The VUP program is divided into two separate programs and the type of program a family qualifies for depends on criteria defined by the Rwandan government including the number of able-bodied adults in the household. The original program is "classic public works" program (cPW), provides cash for manual labor, and a newer "expanded public works" (ePW) program provides cash for typically lighter labor and also provides access to livestock.

Rwanda has 11.9 million inhabitants, approximately 53% under 19 years old. Rwanda has made great progress in addressing infant and child mortality reducing it from 196/1,000 in 2000 to 50/1,000 in 2017. However, deficits remain including a high under-five stunting rate (38%). The country has faced compounded adversities including the 1994 genocide, the HIV/AIDS epidemic and extreme poverty. Although since the year 2000 the real GDP growth has averaged 8% per year, 39.1% of Rwandans live below the poverty line. These adverse factors create a series of challenges, including deficits in nutrition and early childhood development (ECD) education and implementation.

Eighty-seven percent of children less than 6 months old are breastfed but stunting rates rise from 18% at 6–8 months of age to 38% at 18–23 months of age and up to 49% in the lowest poverty quintile. Dietary diversity is limited in Rwanda, especially amongst poor children. In the early years, the need for nutrients to fuel brain development and physical growth is particularly high. A UNICEF/Rwanda Ministry of Health study found that a very limited diet for children—in terms of variety and nutritional content—is common in Rwanda, due both to poverty and poor feeding practices. The study found that a high percentage of children in the poorest households did not receive three meals a day and only 17% of children aged 6–23 months have a minimally acceptable diet. In addition, early childhood development knowledge, attitudes and practices of Rwandan parents are limited. A 2014 UNICEF study found that only 12% of primary caregivers for 2–3-year-old children engaged in three or more activities to promote learning or school readiness (e.g., singing songs, telling stories, teaching the child something new or looking at

pictures in books and magazines with the child). The ECD problem is compounded by the fact that violent discipline is prevalent. An Early Childhood Development and Family Services UNICEF baseline evaluation in 20 sites in Rwanda showed that 19.8% of children 0–11 months old and 80.7% of children 24–36 months old are exposed to some form of violent discipline; 34.4% of caregivers believe that physical punishment is necessary to raise a child well. In addition, men are often not significantly involved in supporting children's early development and decision making is not equally shared by men and women.

Specific Aims

The proposed study will investigate the effectiveness of Sugira Muryango, a family home-visiting caregiver coaching model intended to promote child development, healthy caregiver-child interactions and reduce caregivers' use of harsh punishment among families living in extreme poverty in Rwanda.

The study's Specific Aims are:

Aim 1. To test the effectiveness of Sugira Muryango in promoting child development (cognitive, socio-emotional, language and physical development), reducing harsh punishment and increasing father engagement in childcare in poverty-affected households participating in a social protection program (the VUP program):

> ***Hypothesis Aim 1 (a):*** Children in families participating in Sugira Muryango will demonstrate improved developmental outcomes. Specifically, children from families who received Sugira Muryango will score better on assessments of cognitive, language, and socioemotional development compared to children in the control group (CG).
>
> ***Hypothesis Aim 1 (b):*** Children in families participating in Sugira Muryango will display improved physical growth compared to children in the control group (CG). Specifically, we expect to see improvement in weight-for-age, weight-for-height, upper middle arm circumference. We do not expect to see large improvements in height-for-age over this relative short time frame.
>
> ***Hypothesis Aim 1 (c):*** Caregivers who participate in Sugira Muryango will report lower levels of family violence (intimate partner violence and harsh discipline) compared to children in the control group (CG).
>
> ***Hypothesis Aim 1 (d):*** Among families with a father present in the child's life, families receiving Sugira Muryango will show greater levels of father engagement in caring for the child compared to children in the control group (CG).

Gaps in Knowledge

Throughout sub-Saharan Africa, integrating scalable, cost-effective ECD and violence prevention programs into poverty-reduction and other social welfare programs has great potential

for promoting ECD and reducing familial violence in culturally diverse, low-resource settings. Data from the Sugira Muryango (SM) trial will support and advance the Rwandan government's vision for comprehensive decentralized ECD to help eradicate poverty and violence against children. The objectives of this research seek to contribute to academia and research institutions' roles and responsibilities outlined in the ECD Policy housed under the Ministry of Gender and Family Promotion (MIGEPROF, 2016).

One of the key research questions of the impact evaluation of SM is whether or not the model is effective in improving children's developmental outcomes including socio-emotional, cognitive and language development as well as nutritional status. The objective is to build the evidence base for incorporating key early years caregiving messages to support healthy child development within already established social protection programming in Rwanda, via impact, process, and costing data. If effective, Rwanda is poised to bring such an intervention to scale at a national level, providing an opportunity to share impact, costing, and process evaluations lessons globally.

How this study addresses the Gaps

Rwanda is facing significant challenges, including chronic malnutrition (stunting), early childhood development education and implementation, neonatal mortality, the quality of education, and prevention of violence against children. However, its strengths, political stability, strong governance, fiscal and administrative decentralization, and zero tolerance for corruption, set it among the prime locations where early childhood education interventions can have an immediate effect on the wellbeing of children and families and a long-term positive effect on the development of children and their ability to reach their full potential and fully contribute to Rwanda's future development. As such, the present study addresses these gaps and challenges by assessing the delivery of SM, a deployment focused intervention to promote ECD and reduce family violence.

B. Study Design

Intervention

Sugira Muryango (SM) is a home-visiting model to support playful caregiving, involving all caregivers of the child with emphasis on father engagement whenever a father is present in the life of the child. SM also aims to improve nutrition, care seeking for child illness, family functioning, positive parent-child relationships, and healthy child development. The intervention builds on five core components: (a) builds playful caregiving skills and improves knowledge of ECD to create a safe, responsive and nourishing environment for the growth of young children with a focus on nutrition, health and hygiene promotion; (b) coaches parents of young children in "serve and return" interactions and playful caregiving; (c) develops a "family narrative" to build

hope and highlight sources of resilience for addressing challenges and reducing risks of violence; (d) strengthens problem-solving skills as well as the navigation of formal and informal community resources; and, (e) builds skills in parental emotion regulation, conflict resolution, and alternatives to harsh punishment (Figure 1).

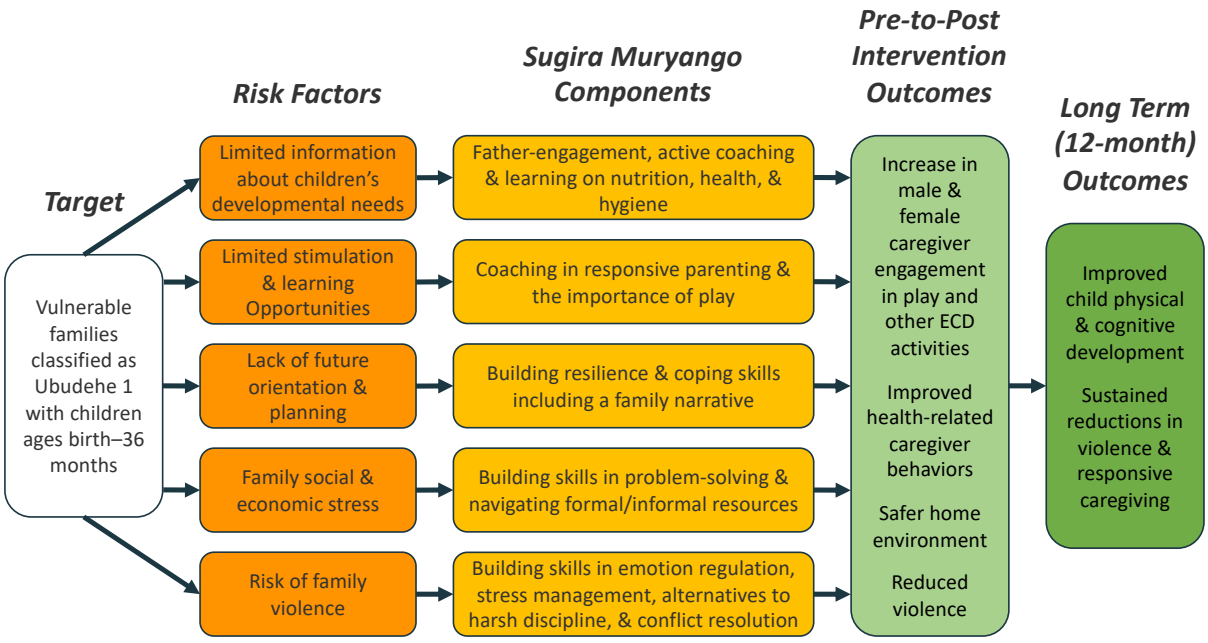


Figure 1. Theory of Change

The components of SM are informed by the WHO's Care for Child Development package and integrated into 12 modules (Table 1), the core intervention, and two booster/follow-up sessions. Each module is delivered in the home of the family and care is taken when planning the home visits to ensure the presence of all primary caregivers of the child, including fathers. Each module takes 60–90 minutes to deliver. The 12 modules are delivered at a pace of one module per week over 3–4 months by well-trained lay interventionists (community-based volunteers referred to as CBVs) who are already embedded in the local community. Primary caregivers participate in the modules in interaction with their child(ren). All home visits include a 15-minute "active play and communication" session where caregivers receive targeted feedback on parent-child interactions. The intervention also focuses on helping families to navigate both formal and informal supports. In this manner, SM connects families to existing resources, harnesses resilience, improves problem-solving skills, and can be delivered through existing service channels, such as the health system or child protection programs. Given its implementation with caregivers who lived during the time of the Rwandan genocide, SM is also trauma-informed by identifying the supports and sources of strength that helped parents make it through difficult times in the past while directing them toward a more hopeful future.

An overview of module session topics is included below:

Table 1. Sugira Muryango Modules, Themes and Goals

Module	Theme	Goals
Welcome Session		• Introduce caregivers to Sugira Muryango • Explain Sugira Muryango goals and structure • Establish a plan for regular meetings together
Module 1	*Family Narrative*	• Discuss the family's priorities and goals • Learn about the family and their children (hopes/goals); Family Narrative • Introduce concepts related to family relationships and ECD and the importance of all caregivers in children's lives.
Module 2	*The importance of responsive caregiving*	• Coach on the importance of diverse play opportunities, responsive caregiving and brain development • Coach positive, responsive parent-child interactions and early stimulation • Expand and strengthen caregivers' repertoire of stimulating activities for their children and toy making • Explain Serve & Return interactions and coach caregivers on this practice
Module 3	*Building early communication skills*	• Coach on the importance of early communication skills and language development • Identify and practice ways to incorporate language learning into play and daily routines • Coach and practice additional techniques to support early speech and language development
Module 4	*The importance of good nutrition*	• Learn about food consumption in the home • Identify nutritional practices that promote child health and growth, including the importance of deworming • Discuss ways to maximize nutrition from available foods
Module 5	*The importance of good hygiene*	• Identify hygiene practices that promote good health • Coach on the obstacles to good hygiene and sanitation
Module 6	*The importance of good health*	• Talk with caregivers about health practices in the home, particularly their response to sick children • Discuss the family's Mutuelle de Sante health insurance coverage • Ensure immunizations are completed/take steps to complete immunizations
Module 7	*Managing the stresses of caregiving and family life*	• Identify and coach caregivers on ways to effectively manage personal and household stresses and frustrations • Explain the importance of adults' consistent emotional self-control for young children

Module 8	*Resolving conflicts in the home*	• Identify and actively coach conflict resolution strategies that promote peace, resilience and well-being in the home • Coach on alternatives to harsh punishment and harmful impact of angry or violent responses to conflict on early childhood development • Coach on the role of positive, responsive caregiving in creating a safe environment for all family members
Module 9	*The important role that everyone plays in raising a baby well*	• Discuss the diverse roles of all family members in supporting early childhood development • Highlight the important role of fathers in raising children • Coach on effective discipline strategies
Module 10	*Good caregiving is better than being born well*	• Coach on the importance of a positive, nurturing parent-child relationship • Coach and practice skills related to responsive, stimulating child-caregiver play and interactions
Module 11	*Making the home a place where a baby's brain can grow*	• Discuss the importance of safety, cleanliness, and support in the home for young children's early learning • Discuss and carry out active coaching on early stimulation activities
Module 12	*With a united family, anything is possible*	• Review program goals and content • Address any remaining questions or concerns • Discuss how the family will use newly learned skills and strategies to promote healthy ECD going forward • Provide family with information on local health/support services

Selection of Community Based lay workers

Community Based lay worker (CBV) coaches are enrolled with assistance from local authorities and leaders in selected study locations. Approximately 118 CBVs will be recruited to deliver SM across the treatment arms. Their caseload of households is approximately 5 households per CBV.

CBVs are selected from the local community using a three-tiered process. First, FXB staff members, referred to as SM expert supervisors, meet with local officials and leaders to discuss SM and introduce the CBV role. Community meetings are held so that community members can learn about the program and express an interest in being considered for the role of CBV. Second, each SM supervisor collates the potential CBVs (either those nominated by village leaders or those who directly express interest) onto a list and conducts a screening interview over the phone using a structured screening guide. Third, the SM supervisor travels to the study districts to conduct an in-depth in-person interview. Once CBVs are selected, they are invited to participate in an intensive three-week training held in their study district.

In addition to the formal selection procedure, CBVs must meet the following inclusion criteria: CBVs selected to deliver the Sugira Muryango intervention must: (a) live in the sector of beneficiary households they will deliver the intervention to, (b) be Rwandan, (c) be aged 18 or older; (d) be able to write, read, and count in Kinyarwanda; (e) be committed to young children and family values; (f) have the required amount of time to carry out the Sugira Muryango intervention with a select number of households, (e) be recommended and approved by local community and authorities.

Training, supervision and fidelity monitoring

Prior to the start of the intervention CBVs undergo extensive three-week training on the SM intervention model taught by Boston based staff and bachelor-level Expert Supervisors who have previous experience with SM. Six bachelor-level Expert Supervisors will continue to supervise the CBVs for a ratio of approximately 1:17.

Supervision and fidelity monitoring are integrated processes intended to support quality of program delivery by the CBVs. All CBVs engage in face-to-face supervision during the first three weeks of the intervention, and each supervisor shadows each CBVs once in the home. Telephone supervision and peer support groups occur weekly throughout the intervention and group supervision is held once a month. During each home visit the CBVs ask for the caregivers' permission to audiotape the home-based sessions. These recordings are reviewed by an expert supervisor who provide targeted feedback to home visitors to strengthen their work on active coaching and father engagement. This fidelity monitoring feedback loop informs the supervision of the CBVs so knowledge and skills pertaining to high quality delivery are supported.

Recruitment and Study Population

The targeted study population (referred to as "beneficiaries") are families with at least one child aged 6–36 months, who are eligible for Rwanda's social protection program Vision 2020 Umurenge Program (VUP) and live in one of the selected districts: Nyanza, Ngoma, and Rubavu.

Family inclusion criteria are (a) living in the Rubavu, Ngoma or Nyanza District of Rwanda. Inclusion criteria for families are: (b) being VUP-eligible (according to the Rwandan government), (c) having at least one child aged 6–36 months living in the home, and (d) having at least one caregiver who is willing to discuss and enhance their caregiving practices by interacting with a home-visiting coach (CBV).

Further caregiver inclusion criteria are: (a) is aged 18 or older and cares for child(ren) and (b) lives in the same household as child(ren). We will enroll both single and dual caregiver families

to reflect population dynamics. Legal guardians may be parents, aunts, uncles, grandparents, or foster parents.

Family exclusion criteria are not meeting inclusion criteria, caregiver having severe cognitive impairments which preclude their ability to speak to the research questions under study, or families or caregivers are in the midst of crisis (e.g., a caregiver with active suicidal attempts or psychosis). Families with ongoing crisis or disability will be referred to appropriate services.

Consent Procedures

During household enrollment, Laterite – the local data collection subcontractor - uses VUP lists to go door to door to assess household eligibility and offer participation in the SM home visiting intervention. If a household is eligible and wants to enroll, the field team consents the primary and secondary caregivers and the VUP head of household if different from the primary and secondary caregiver. The primary caregiver also provides consent for the child ages 6–36 months. If a caregiver is absent, the team make a plan to come back either later in the day or the next day. The field team gives each home a week for this to be completed before replacing the home. Given that most Ubudehe 1 individuals (those eligible for VUP public works) are illiterate, the consents are orally communicated to the caregivers. The caregivers either sign their name or make a mark (thumbprint) on the consent form. The Laterite field team carries additional copies of the consent in case the family would like to keep a copy. The consent forms are stored in a portable lock box until they are transferred to Kigali and stored in locking file cabinets at the FXB-Rwanda office. Study participants are reconsented prior to each follow-up data collection timepoint.

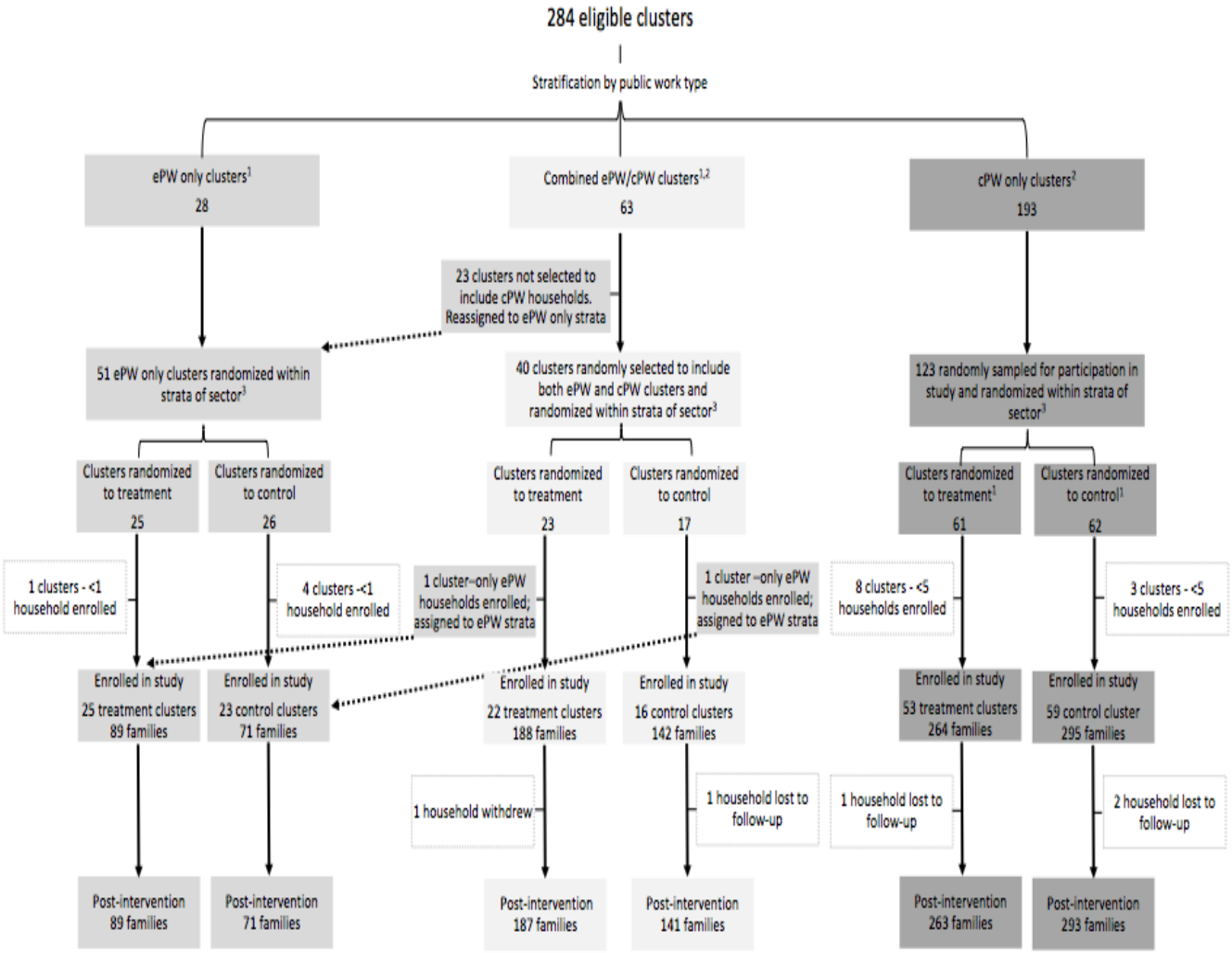
Randomization Plan

Families' participation in the VUP and selection into cPW and ePW is determined by governmental policies and was not under the control of the research team. Lists of families participating in the VUP program were obtained from government staff in each district.

Nonoverlapping, geographically defined clusters were created comprising at least 30 families participating in the cPW program or ten families participating in the ePW program, with some clusters containing both ≥30 cPW and ≥10 ePW households. Clusters were formed from one or more contiguous villages such that one CBV could provide services to all participating families in the cluster. Villages within the same cluster were selected to be as close to each other and as far apart from other clusters as possible. Due to the relative scarcity of the ePW families, 100% of clusters containing at least 10 ePW families were sampled for participation in the study. Clusters which contained cPW families (including combined clusters containing ePW families) were randomly sampled for inclusion into our study until we reached our target sample size of ≥1,040 households. Randomization was performed by Laterite and occurred at the cluster level within strata defined by public works type (ePW only, combined ePW/cPW, and cPW only) and

geographic sector. Within strata, clusters were assigned random numbers and placed on a ranked list. The first half of clusters on the randomly ranked list were assigned to treatment. In case of an uneven number of clusters per strata, a lottery was used to round the number assigned to treatment up or down. After assignment of the cluster, households were contacted by the data collection contractor and invited to participate in the study. Clusters were retained if at least five families in the cPW strata or at least one family in the ePW strata enrolled and had at least one child aged 6–36 months. We retained 48 ePW-only clusters, 38 ePW/cPW clusters, and 112 cPW-only clusters (Cluster sampling strategy, Figure 2).

Neither the families nor the enumerators who conducted the assessments knew about a family's assignment to treatment versus control before they had completed the baseline assessment. All caregivers gave written informed consent for themselves and their eligible children ages 6–36 months.



cPW = expanded public works; cPW = classic public works; [1]Clusters had to include ≥10 ePW households to be eligible as an ePW cluster; [2]Cluster had to include ≥ 30 ePW households to be eligible as a cPW cluster; [3]Randomization within strata of Public Works x Sector did not necessarily guarantee even numbers in the treatment and control arms

Figure 2. Cluster sampling strategy and flow Chart of Participants in the *Sugira Muryango* Open Trial.

Note: Although each cluster had a 50% chance of being assigned to receive treatment, we were not guaranteed an equal number of treatment and control clusters because randomization occurred within relatively small strata that sometimes contained an odd number of clusters.

Study Visits

Data collection occurs at three timepoints referred to as baseline, midline (4 months post intervention) and endline (12 months after midline). All assessments are carried out in the homes of the families or at a central point in the village.

At each time point, the caregiver who states that he or she knows the child best —most often the biological mother—provides reports on child development, health, feeding practices. This "primary caregiver" also participates in the assessment of caregiver-child interactions and provides information about the household including family composition, economic status, household assets, social protection, and finances. Both the primary caregiver, their intimate partner or other secondary caregivers living in the households (e.g. grandparents, adoptive parents, aunts and uncles) respond to a battery of questionnaires that cover aspects of caregiver mental health, trauma exposures, daily hardships, family functioning, decision making, alcohol consumption, and intimate partner violence. Measures were drawn from previous pilot and research studies in Rwanda (Betancourt et al. 2018) and followed a rigorous translation protocol, including forward- and back-translation from English to Kinyarwanda.

The order in which the different surveys (The Report on the Child, the Caregiver Report on the Household, and the Caregiver Report on Self) are administered depends on whether the child is able to be assessed. If the child is present and awake, the Report on the Child is administered first, then the other surveys follow. If the child is asleep or not available, the enumerator starts with the caregiver surveys instead. Due to the sensitive nature of the intimate partner violence questions in the Caregiver Report on Self, this survey is conducted only by enumerators of the same gender as the respondent and in as private a location as possible. In the case that a survey is not completed during the first visit—because there is a mismatch in the genders of the enumerator and respondent, for example—an follow up appointment is communicated to the household immediately and an enumerator will return at the earliest opportunity.

Participants attend appointments for the child assessment on the Malawi Development Assessment Tool (MDAT) and anthropometric measurements at central location. A community health worker selects a safe place—such as an office, a church, or school—to serve as the data collection site. Participants are notified in advance regarding the location and timing of these assessments.

Enumerator training for baseline data collection occurred from January 29, 2018 to February 9, 2018. Training for surveys completed at the household included an overview of the study's objectives and methodology, sampling and replacement strategies, research ethics, the field team's responsibilities, adverse event reporting and the risk of harm protocol, and a thorough review of the survey instruments. A separate training occurred for those enumerators overseeing the MDAT and collection of Anthropometrics. After the training, an additional day of piloting all measures occurred in the field. A three-day refresher training occurred from April 18 to April 20, 2018. Prior to collection of midline data, a refresher training was held between August 1 and August 8, 2018. As there were no major changes to household survey questions, no pilot was done in advance of midline data collection. A three-day refresher training was also held for the MDAT enumerators and a midline pilot of the MDAT occurred to test for interrater reliability.

Key measures collected at baseline (wave 1), midline (wave 2) and endline (wave 3)

Treatment status variable
Household treatment status is considered the main predictor of all outcomes.

Public Works program
The household's public work (VUP) status is considered an important design-variable.

Child anthropometrics
Children's physical growth status is used as a proxy for their nutritional status and assessed using measures of standardized height-for-age (HAZ), standardized weight-for-age (WAZ), standardized weight-for-height (WHZ), and middle upper arm circumference (MUAC). Height is measured in centimeters to the nearest 0.1 cm. Weight is measured in kilograms to the nearest 1 gram. Middle upper arm circumference is measured in centimeters to the nearest 0.1 cm. Standardization is done using software provided by the World Health Organization. Outcomes related to children's HAZ, WAZ, WHZ and MUAC are treated as continuous.

Malawi Developmental Assessment Tool
Child development across domain of gross motor, fine motor, language and socio-emotional development is assessed using the Malawi Developmental Assessment Tool (MDAT) tool.

MDAT was developed for use in Malawi with the aim to develop a culturally appropriate tool for use in rural Africa (Gladstone et al. 2010). MDAT is an observational task-based assessment where a trained rater assesses the child's ability to perform different tasks. The MDAT has been tested in several African countries where it has shown good reliability and validity. We calculate a standardized continuous score for each domain Fine motor, Gross motor, Language, and Socioemotional using standardizations provided by the MDAT team.

Ages and Stages Questionnaire

The Ages and Stages Questionnaire 3 (ASQ-3) uses a series of parent-completed questionnaires that are designed to screen the developmental performance of children in the areas of gross motor skills, fine motor skills, communication, and problem solving (Squires & Bricker, 2009). We administer the ASQ-3 through oral interviews with the caregiver who states that the know the child best ("the primary caregiver"). The ASQ-3 is scored using standard guidelines for standardization resulting in create binary outcome variables for each domain indicating whether or not the child showed "developmental delay" vs. "no developmental delay".

Violent disciple

Children's exposure to violent and nonviolent discipline was assessed using the MICS Child Development and Child Disciplinary modules. Exposure to violent disciplinary practices was defined as having experienced as a cumulative score of the number of forms of violent discipline

Intimate Partner Violence

Intimate partner violence among parents who are currently married, cohabitating, and/or in a relationship was assessed using an adapted and abbreviated version of the revised Conflict Tactics Scale (CTS-2: Strauss, 2004) previously used in Rwanda (Chaudhury et al. 2016). The CTS-2 assesses caregivers' self-reported victimization and perpetration of emotional neglect/abuse, physical assault, and sexual assault. Given the short interval between the baseline and midline data collection, the scale was adapted so that caregivers report on events that occurred within the last three months and ever (as opposed to 12 months in the original scale). For events that had happened, caregivers further report on the frequency of that event using the following scale: 0="not at all"; 1= "once:"; 2 ="Twice"; 3 = "3-5 times"; 4 = "6-10 times" ; 5 = "11-20 times"; 6 = ">20 times." Overall, the same events are asked for victimization and perpetration with the exception that two additional sexual assault questions were asked about for victimization (detailed below). The following examples refer to victimization. After the questions about victimizations, caregivers were asked similar question for perpetration stated as "You [left you or forgot] your partner" etc.

Two items assess emotional abuse or neglect: "Your partner used abusing language toward you" and "She/he left you or forgot you." Seven items assess physical abuse : "She/he shoved you,"

"She/he grabbed you," "She/he hit you," "She/he used a knife or any other weapon," "She/he hit you with an object that could cause injuries," "She/he pushed you into a wall," "She/he kicked you." Sexual abuse is assessed using three items for victimization: "She/he forced you have sexual intercourse when you didn't want," "Have you ever had intercourse with him/her because you were afraid of what she/could do if you refused" and "Has he/she ever forced you to take part into sexual activities that you found shameful or disrespectful?" For perpetration only one item assessed sexual violence "You used forceful means to be able to have sexual intercourses with your partner." In line with the recommendation for community samples we use a count of events as the outcome variable (Strauss, 2004). We counted only physical and sexual abuse events given the milder nature of the emotional abuse/neglect items. An alternative approach is to recode the frequency score of violent events in into midpoint scores as described by the creators of the CTS2. This, however, is not recommended since this score has been found to be extremely skewed (Strauss, 2004).

Father engagement in childcare

Father engagement in childcare is measured using the following item "Does the father of the child spends some time every day caring for the child?" which is reported by the primary caregiver of the child (usually the mother). The response is recorded as a binary "yes/no" outcome variable.

Other procedures at Midline (4 months post-intervention)

Qualitative Interviews

A subsample of 40 caregivers who receive the intervention participate in qualitative interviews to assess their experience of the program using an IRB approved interview protocol. A subsample of N=40 CBVs participate in qualitative interviews to assess their experience of delivering the program.

Quantitative data collection

Post intervention, CBVs complete a fidelity, acceptability, and utility measure to gauge how effectively they felt the training prepared them to deliver the intervention, as well as how often they used the skills learned during their professional and personal life. This measure was adapted from Aarons, Ehrhard et al. (2015) to assess the implementation science behind training lay workers to deliver technical interventions.

Quality Control

Field Supervision

Throughout data collection, the field supervisor for the surveys at the household chooses one sub-team to accompany each day to confirm that interviews are conducted in the right households, protocols are followed, and ethical considerations are met. A chain of command allows enumerators to raise issues to the field coordinator, who then reports them immediately to the field supervisor, who communicates them to the data manager for the final decision.

At the end of each day, the field supervisors update the log of surveys completed and issues to be cleaned in the data and compile a daily field report for the data manager. The teams for both the surveys at the household and the MDAT and anthropometric assessments also attend a daily debrief session led by the senior field supervisor and data manager. During the debrief, the team discusses issues from the day's surveys, and the data manager provides solutions and communicates any changes in the survey. Before heading to the field each morning, the team meets with the data manager who updates the list of households with any replacements from the previous day, reviews the schedule, and addresses quality concerns from the data collected in the previous days.

Real-time Completion Tracking
Laterite uses SurveyCTO's feature of real-time publishing of form submission data into Google Sheets to track the progress of data collection compared to the targets in the field plan. Dashboards are created to track the completion of the required surveys for each household (household completion status) and then the completion of surveys for all households in a cluster (cluster completion status.) The cluster completion status is shared with the Boston College team in real time, which allowed them to start the intervention in the households of completed clusters as soon as possible. The Google Sheets were also used to track replacements of caregivers and households and flag duplicate surveys.

Routine Monitoring
Laterite uses proprietary audit algorithms to review survey metadata to flag unusual submissions such as those with changes of location, early or late starting times, comparatively short or long durations, and simultaneous submissions. Concerning surveys are flagged to the Data Manager and Senior Field Supervisors for further investigation. In addition to the real-time monitoring using Google Sheets, Laterite routinely monitors the cleaned data for survey duplication, household and caregiver replacement, household dropout, child dropout and mortality, and household survey completion. Laterite reports this information to Boston College in the Weekly Data Collection Reports.

Audio Audits
Audio from a sample of all surveys conducted at the household are automatically recorded using the in-built recording feature of SurveyCTO. The questions that triggered the audit are pre-selected and recordings are reviewed to assess whether (i) the interviews actually took place; (ii)

enumerators are following proper interview procedures such as: explaining confidentiality and sticking to the script while asking sensitive questions; (iii) enumerators are asking questions with a respectful tone and without pushing the respondent or leading them towards a certain response; and (iv) explaining to the respondents that they could opt out of taking part (or answering questions in some parts of) in the survey. A random sample of 5% of all surveys are recorded for each question.

No audio audits are set for the MDAT and anthropometrics surveys since the proper administration of these assessments relies less on enumerator-participant dialogue. These surveys are instead administered under supervised conditions and enumerator behavior is closely monitored.

Throughout data collection, Kinyarwanda-speaking data auditors review the audio recordings to confirm that there is dialogue between the enumerator and the respondents (i.e., that responses are not entered without actually asking the questions) and that enumerators are following proper interview procedures like explaining confidentiality and ensuring privacy for sensitive questions and asking questions in a respectful tone without pushing the respondent or leading them toward a certain response. Issues identified by the auditors are recorded and relayed to the data manager for proper follow-up and resolution with the field supervisors.

C. Data Management

Data Collection

Questionnaires and child assessments are administered verbally in Kinyarwanda. Data are entered into SurveyCTO directly on Android tablets by trained independent local enumerators blinded to intervention status. Once the data are initially processed and de-identified, they are uploaded to a secure, password-protected Box folder. The data are further backed up to a secure server at Boston College.

Database

Data collection includes the collection of quantitative data regarding intervention effectiveness, dissemination and implementation (D&I) data regarding key domains of implementation science (e.g. acceptability, appropriateness, feasibility), qualitative data via face-to-face key informant interviews, and fidelity data collected throughout intervention delivery.

Quantitative study data and D&I survey data are collected using SurveyCTO, a secure mobile data collection platform that can be used offline. The SurveyCTO mobile application, referred to as SurveyCTO Collect, allows an enumerator to collect participant data using a password protected mobile phone or tablet, which is then uploaded to the secure SurveyCTO server. SurveyCTO has a built-in data monitoring and visualization tool that allows Field Coordinators

and Field Supervisors, who are based in the field during data collection, to monitor the uploading of study data as well as any inconsistencies in the data. Data will be regularly uploaded from tablets to the password protected SurveyCTO cloud-based server. Whenever form data is transmitted via a 3G or other internet network, it is encrypted in transit.

Qualitative key informant interviews are collected from primary caregivers (SM beneficiaries), CBVs, local and national government officials and other ECD and nutrition stakeholders such as International and local nonprofit organization and and international funding agencies. The aim of these interviews is to identify best practices and quality of early childhood and nutrition services delivery in Rwanda. A semi-structured interview guide is used to guide the qualitative interviews.

D. Data Analysis

Primary Outcomes

See Measures section for detailed information about the variables.

The table below outlines study primary predictors, control variables and moderators (possibly mediators if relevant).

	Questionnaire / Scale
PRIMARY PREDICTOR VARIABLES	
Treatment status [Treatment vs. Not]	
OTHER PREDICTOR VARIABLES [design variables]	
Cluster	
VUP-status [classic public works vs. expanded public works]	Government records
OUTCOME VARIABLES	
Malawi Developmental Assessment Tool (MDAT)	
Gross motor [standardized score]	MDAT
Fine motor [standardized score]	MDAT
Language [standardized score]	MDAT
Social [standardized score]	MDAT
Ages and Stages Questionnaire (ASQ-3)	
Gross motor [delayed/not delayed]	ASQ-3
Fine motor [delayed/not delayed]	ASQ-3
Communication [delayed/not delayed]	ASQ-3

Personal social [delayed/not delayed]	ASQ-3
Problem solving [delayed/not delayed]	ASQ-3
Anthropometrics	
Height-for-age (HAZ)	Anthropometrics
Weight-for-age (WAZ)	Anthropometrics
Weight-for-height (WHZ)	Anthropometrics
Middle upper arm circumference (MUAC)	Anthropometrics
"Father engagement "	
" father spends time every day caring for the child"	HOME
Family Violence	
Violent discipline [cumulative score, 0-7]	MICS-5 Hash discipline module
Intimate partner violence	
Victimization physical and sexual abuse past 3 months [cumulative score 0-7]*	Conflict Tactics Scale- CTS2
Perpetration physical and sexual abuse past 3 months [cumulative score 0-7]**	Conflict Tactics Scale- CTS2

*Among cohabitating or married mothers only
**Among cohabitating or married fathers only
***Among cohabitating or married caregivers only

Table 2. Study predictor and outcome variables.

Data Analysis Plan

Standard descriptive statistics will be used to describe the study sample. Frequencies and percentages will be reported for categorical variables and mean, median, standard deviation, interquartile range, maximum and minimum values will be reported for continuous variables. Anthropometric data will be cleaned for extreme out-of-range values. Scales will be created from individual items according to scoring instructions for each scale. All scales will be evaluated for psychometric properties. Visual inspection of histograms and other plots will be used to identify outliers.

Continuous Outcomes MDAT, WAZ, MUAC, WHZ: The goal is to determine the efficacy of SM at 12 months from study initiation. Measurements of all outcome variables (see Table 2 and Section B), are taken at baseline, midline (4 months post baseline), and endline (12 months post midline). Subjects are randomized (Section B) into a treatment group (SM) and CG. Effectiveness will be determined if the slope of the response variable for the SM group is significantly different than the slope of the CG group. To test this hypothesis, we will fit a linear

mixed effect model with a MDAT subscale, for example, as the response variable and the primary predictors are treatment group (SM vs CG), time and their two-way interaction included as fixed effects. Mixed effect models also known as hierarchical linear models (HLM) or multi-=level models are a flexible tool for analyzing associations and changes over time in longitudinal studies when there are clusters of correlated data in the outcome variable. This design has three levels of nesting, families are measured within measurement waves, and measurement waves are nested within randomization cluster or region. Since, we expect region and time level effects subject-specific slopes and intercepts will be modeled as random effects nested within randomization cluster also modeled as a random effect. Additionally, there may be due to VUP program (recall we stratify the randomization of clusters within public works cluster). Type of VUP program will be model as a fixed effect. A mathematical representation of this model is:

$$Y_{ijkm} = B_0 + b_{0i} + B_1*TreatmentGroup + B_2*Time + b_{1i}*Time + B_3*Time*Treatment\ group + b_{2k} + b_{3k}*time + B_{4m}*PubicWorksGroup + e$$

Where, Y_{ijkm} represents the value of the i_{th} family $(i = 1 \ldots I)$, at the j_{th} time point $(j=1, \ldots 3)$, in the k_{th} cluster $(k = 1, \ldots 200)$ and m is the VUP indicator with levels expanded public works, classic public works and mixed. The set of b_{0i} variables represent subject-specific intercepts and are assumed $b_{0i} \sim N(0, \sigma_1)$. The b_{i1} are subject specific slopes and assumed $b_{1i} \sim N(0, \sigma_{slopes})$, the b_{2k} are randomization cluster intercepts and assumed $b_{2k} \sim N(0, \sigma_{region})$ and the $b3k$ ~are region specific slopes and assumed $b3k \sim N(0, \sigma_{sregion})$. The value e represents residual variation and is assumed $e \sim N(0,\sigma)$. Efficacy will be based on the likelihood ratio test for inclusion of the Time*Treatment interaction (null hypothesis $B_3 = 0$). Random effects will be subject to a limited amount of model building as we would like to simplify the model where appropriate. For example, we would want to test if we can assume a common cluster slope versus a unique slope for each cluster. Cluster and subject-specific effects will be dropped from the model based on a Likelihood Ratio Test (LRT) for their inclusion. The estimation procedure while model building random effects will be Restricted Maximum Likelihood (REML). The primary hypothesis test which is based on the inclusion of a fixed effect will require models estimated by maximum likelihood. Because of randomization we have controlled in theory for all variables observed and unobserved, however treatment group imbalance may still be present. We explore treatment group characteristics for in-balance. However, previous work has showed baseline differences to be minimal.

Model assumptions and alternative tactics: Model assumptions will be verified. In particular we will examine the distributions of the residuals to verify that normality assumptions hold. Residual plots will reveal violations of the normality assumptions. Panel plots will be examined to look at functional shape assumptions within subject. In the case of non-normality of outcomes, we will consider response transformations or the use of models that do not require the normality assumptions including the class of generalized linear mixed effect models. If the linearity

assumption does not hold across time (baseline, 3 months, and 12 months) we can treat time as a categorical variable to examine differences between the treatment groups at individual time points focusing on the 12-month group differences adjusted for baseline. If examination of panel plots reveals a large amount of heterogeneity among trajectories of the outcomes by subjects we can use Growth Mixture Models which first identify clusters of similar trajectories from which we can test for treatment group differences in terms of membership in the clusters of similar trajectories.

Outliers will be flagged through visual inspection and standardized residual values greater than 3. A sensitivity analysis will be done to determine if outlier removal changes the coefficient values. If we find that outliers are having an oversized effect on the estimated coefficient values we will give preference to the models where the outliers are removed. However, we will present both results.

Some responses are scales or derivations of scales in particular the MDAT and the ASQ. We will examine the psychometric performance of these scales. We are especially concerned with internal consistency and item reliability. If these scales do not show at a minimal amount of internal consistency (Cronbach's Alpha > 0.6) these items will not be considered for analysis. If particular items are an issue they may be dropped from the scale. Research, in areas like Rwanda suffer because standard scales have not been validated on these populations.

Binary Outcome ASQ, shared decision making, father engagement: The analysis for these outcomes will be similar to the analyses above with accommodations being made for the fact the ASQ is a binary outcome. In particular, we will have to make use of mixed effect logistic regression models a form of generalized linear mixed models (GLMM) with a binomial distribution and a logistic link. Outside of accounting for the correct distribution of the response variable the analysis will follow the methods outlined in the previous MDAT section.

Model Assumptions and Alternative Tactics: We will examine Pearson residuals to assess goodness of fit. Alternative methods include the use of quasi-binomial mixed effects regression (in the case of over dispersion). However, such models do not have a full likelihood and instead of LRT test we will have to rely on the Wald-type type test for regression coefficients.

In fitting logistic mixed effect regression models, it is possible that we may run into numerical difficulties in computing a likelihood ratio test, so score type tests may be needed. This is more likely to occur in the case of testing for variance components.

Count Outcomes, harsh discipline, physical and sexual violence: These metrics count types of harsh discipline action or physical and sexual abuse as suffered or inflicted. Our experience with these scales is that there are more values of zero than what would be predicted by a standard

Poisson distribution. Our analysis will make use of Zero Inflated Poisson (ZIP) models (Everitt, & Hothorn, 2005; Long, 1997). These models belong to the more general class of mixture models. In particular they model the binomial probability of a count of zero versus a count greater than zero and the Poisson probability for counts greater than zero. In particular we will fit mixed effect models as described in the continuous variable sections but under the ZIP distribution assumption.

A key difference between a mixture model and more standard probability distributions is the fixed effects structure and tests for significance can vary between the two parts (binomial and Poisson). Likelihood ratio testing for the Time*Treatment group interaction will be performed for both components of the mixture model. A significant result of the LRT test for the binomial portion would indicate that the probability of no violence (inflicted or suffered) differed between the treatment groups. A significant result of the LRT on the Poisson portion of the ZIP model would indicate that among those who inflicted or suffered harsh discipline or violence the mean amount of violence differed between the treatment groups. The two portions are contained in a single likelihood function and parameter estimation is done simultaneously.

Model Assumptions and Alternative Tactics: The key assumption in these models is the ZIP distribution of the response variables. We will visually inspect the distribution of the outcome variables to verify that we are fitting the correct model. Alternative distributions could be Poisson, Quasi-Poisson, Binomial (data dichotomized due to lack of variability), or Negative Binomial. These are non-mixture models and modeling will follow the process outlined in the continuous outcomes section.

Secondary Outcomes: Stunting measured by Height-for-Age is an important outcome. However, we do not expect to see significant changes in this variable in the 12-month timespan. We do expect to see a reduction in stunting even if it does not achieve statistical significance. Using the models described in data analysis plan we can extract the coefficient for the time by treatment group interaction as an estimate of the size of the effect of SM on stunting. Effect sizes of other secondary outcomes will be reported similarly.

Missing Data

We will apply the ITT principle to all of our analyses. Unless a subject dies, the ITT convention is to keep them in the study; HLM handles this well since the linearity assumption allows us to estimate subject-specific parameters so long as we have at least two data points. The method of mitigating bias due to missingness known as full-information maximum likelihood estimation (FIMLE) can be problematic in mixed effects linear models. Principled methods of multiple imputation such as chained equations will be employed.

Power and Sample Size

Assuming the linear mixed effect model described above. The following values were used to calculate statistical power. We assumed a total of 200 clusters with 5 families per cluster. We assume a linear relationship over time. The significance level is set at 0.05. The statistical test was the likelihood ratio test for the inclusion of the time by treatment group interaction. The ICC was set at 0.10. A variable for public works type which was categorical with 3 levels was included in the model. Within public works type a ratio 1:1:2 for expanded:both:classic was assumed. Variability was set to one so that effect sizes may interpreted as "Cohen-Like" effect sizes. We found that we have at least 90% power to detect an effect size of 0.25 and at least 80% power to detect an effect size of 0.22. GLIMMPSE software was used for these calculations.

E. Safety & Adverse Events

Risks to Subjects

The following risks were identified and submitted to the Boston College IRB:

1. Participation in the intervention may lead adults to confront stressful personal or family issues or generate concern about social harm resulting from loss of confidentiality, which may elicit distress in some participants.
2. In some cases, fatigue from the psychosocial assessment administered pre/post-intervention and at 12-months follow-up may occur.
3. There is a small risk of loss of confidentiality.
4. Participation in the intervention may create issues within the community if a household receiving the SM intervention is viewed as receiving additional services or monetary support.

We consider most risks associated with participation in the study to be unlikely. Two previous pilots of SM have been carried out with vulnerable families in Rwanda. Previous study iterations have demonstrated that processes related to confidentiality are upheld throughout the entire study. Further, no participants have indicated distress or fatigue with the assessments. Weekly field reports from the Laterite enumerators have not indicated any issue with the assessments, length of time to administer, issues with confidentiality, or distress caused.

There are other risks we consider to be likely: intimate partner violence in the home and tensions with neighbors due to participation in the intervention. The CBVs have received comprehensive training on recognizing intimate partner violence in the home as well as how to activate our risk of harm protocol to ensure we assess the harm and make the necessary referrals. The other risk related to community tensions due to participation have resulted in several study changes. First, in the event that an issue arises with a neighbor, the CBV immediately alerts their supervisor and the village leader or elder to better understand the source of the issue. During this time, the intervention is paused. The family is then given the option to conduct the intervention

offsite at the local government offices at the cell level. As a preemptive measure, all supervisors have traveled to the field to meet with the village leader and Cell official to further communicate the SM as a coaching intervention and not an intervention where goods or money are provided to families.

Monitoring

A Risk of Harm decision tree, along with flagged questions from the assessment battery, provides a structured protocol for monitoring and responding to adverse effects on participants. Interventionist supervisors have daily phone check-ins with the interventionists as well as weekly face-to-face meetings. These check-ins provide an opportunity for monitoring families in the study and ensuring all study risks are addressed appropriately. Further, these check-ins allow the supervisor to anticipate any potential risks and work with the interventionist to initiate support or a referral to mitigate any potential risks.

Stopping Procedures if relevant.

The intervention will be paused or stopped if a participant is experiencing active psychosis or is experiencing harm as a result of participation in the intervention. While these instances are expected to be rare, formal procedures are in place. If an adverse event or risk of harm is detected, a study supervisor works with the CBV to collect as much information on the issue, and then conducts an in-person assessment of the situation. Depending on the case, this assessment may include discussion with local CHWs and village leaders to gain further information about the participant or family.

F. Study Limitations

A primary study limitation is the reliance on caregiver-reported measures, which could suffer from differential bias because parents who are exposed to the intervention may be more knowledgeable about or more pressured to provide a socially desirable answer. However, the quality monitoring approaches utilized by study enumerators (described in detail above) provide important checks for ensuring the data are collected as intended and that enumerators are not engaging in any sort of behaviors to coerce specific answers from caregivers. A second study limitation regards the reliance on western created measures to assess primary and secondary study outcomes. Extensive work was done to refine and adapt measures to fit the Rwandan context with assessment questions forward and back translated into Kinyarwanda following best practices (Van Ommeren et al., 1999) though two pilot studies (Betancourt et al. 2017 and other forth coming manuscripts). Further, psychometric analysis of study measures assessing primary and secondary study outcomes indicate strong reliability and validity.

G. Compensation

Incentivization of CBVs

CBVs receive several types of incentives given their role in delivering SM. The incentive rate was established based on learning from what other similar programs require for a volunteer coach as well as considerations regarding the CBV scope of work. Calculated at a per household rate, CBVs will receive 1,200 Rwandan Francs each week per household in their caseload to cover cell phone airtime, transportation, and the family visit. CBVs also receive 2,000 Rwandan Francs to cover transportation for weekly peer support groups over the course of 12 weeks. To cover their participation in a three-week intensive training, CBVs receive an additional 75,000 Rwandan Francs. Non-monetary incentives include an umbrella, branded t-shirt, mobile phone, lockbox for sensitive materials (intervention manual, intervention workbook, notes from supervision), rain boots, a branded bag, a branded identification badge, and materials needed to deliver the intervention.

Incentivization of Household Beneficiaries

Household beneficiaries will receive 5000 Rwandan Francs for each timepoint of data collection (baseline, post-intervention, and 12-months follow-up). Households selected to participate in qualitative interviews will receive an additional 3000 Rwandan Francs for each qualitative interview. Households selected to participate in the D&I measures will receive an additional 3000 Rwandan Francs for each time point (pre- and post-intervention). Families are reimbursed in cash once they have completed the assessment or interview.

H. Provisions for vulnerable subjects

All household beneficiaries in the study have an Ubudehe 1 poverty categorization. As such, all study procedures are created and implemented with provisions for vulnerable participants. Specifically, risk of harm/adverse event procedures are in place to identify any risk of harm situations related or not related to participation in the study. Further, funds are available if any participant requires a referral or transfer to a higher level of care, such as transport to and treatment at a district hospital, for example.

I. Data archiving and dissemination

Results from the trial will be published in peer-reviewed journal articles and presented at high level conferences. A formal dissemination event involving study funders and stakeholders will be held in Kigali, Rwanda. Study quantitative data will also be uploaded to the World Bank Microdata Catalog as part of study funding deliverables.

References

Betancourt, T. S., Franchett, E., Kirk, C. M., Brennan, R. T., Rawlings, L., Wilson, B., ... & Ukundineza, C. (2018). Integrating social protection and early childhood development: open trial of a family home-visiting intervention, Sugira Muryango. *Early Child Development and Care*, 1-17.

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 for the Lancet Early Childhood Development Series Steering Committee. (2017). Nurturing care: promoting early childhood development. *The Lancet*, 389(10064), 91-102.

Chaudhury, S., Brown, F. L., Kirk, C. M., Mukunzi, S., Nyirandagijimana, B., Mukandanga, J., ... & Betancourt, T. S. (2016). Exploring the potential of a family-based prevention intervention to reduce alcohol use and violence within HIV-affected families in Rwanda. *AIDS care*, *28*(sup2), 118-129.
DOI: 10.1080/09540121.2016.1176686

Engle, P. L., Fernald, L. C., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., ... & Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, *378*(9799), 1339-1353.

Everitt, B. S., & Hothorn, T. A. (2005). *Handbook of Statistical Analyses Using R*. Boca Raton, FL: Taylor and Francis.

Gladstone, M., Lancaster, G. A., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N. R., & Smyth, R. L. (2010). The Malawi Developmental Assessment Tool (MDAT): the

creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS medicine*, *7*(5), e1000273.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Ministry of Gender and Family Promotion. (2016). Early Childhood Development Policy. Kigali, Rwanda: Ministry of Gender and Family Promotion.

Richter, L. M., Daelmans, B., Lombardi, J., Heymann, J., Boo, F. L., Behrman, J. R., ... & Bhutta, Z. A. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet*, *389*(10064), 103-118.

Shonkoff, J. P., & Fisher, P. A. (2013). Rethinking evidence-based practice and two-generation programs to create the future of early childhood policy. *Development and Psychopathology*, *25*(4pt2), 1635-1653.

Straus, M. A. (2004). Scoring the CTS2 and CTSPC. *Family Research Laboratory, University of New Hampshire.(On-line)*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.295.2407&rep=rep1&type=pdf

Straus, M. A., & Douglas, E. M. (2004). A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence and Victims*, *19*(5), 507-520.

Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires®, Third Edition (ASQ®-3): A Parent-Completed Child Monitoring System*. Baltimore: Paul H. Brookes Publishing Co., Inc.

Van Ommeren, M., Sharma, B., Thapa, S., Makaju, R., Prasain, D., Bhattarai, R. & de Jong, J. (1999). Preparing instruments for transcultural research: Use of the translation monitoring form with Nepali-speaking Bhutanese refugees. *Transcultural Psychiatry, 36*, 285-301. doi.org/10.1177/136346159903600304