Understanding longer-term effects of a father-engaged, play-based home visiting intervention to promote early childhood development and prevent violence in Rwanda: The Sugira Muryango Longitudinal & Spillover Study

(A follow-up of "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: May 25, 2022

1. Background

This study is the culmination of several years of research in Rwanda by the investigators. Initial research led by Dr. Betancourt and Dr. Sezibera explored the local perceptions of mental health problems among HIV/AIDS affected youth and adversely affected households to identify coping strategies and sources of support relevant to the Rwanda context and culture. Building on this work the team used findings to guide the development and adaption of quantitative measures to assess prevalence of mental health problems in children and caregivers. In later years, the team developed an adapted Mental Health Intervention for HIV/AIDS-Affected Children, which was piloted to test feasibility and acceptability of the intervention. This intervention, referred to as *Sugira Muryango*, was further adapted in recent years to be focused on early childhood development in general. The Early Childhood version of *Sugira Muryango* is a family-based, home-visiting intervention targeted at early childhood development and implemented with families living in extreme poverty in three districts of Rwanda. This version of *Sugira Muryango* was first tested in two small pilot studies (Betancourt et al. 2018; Barnhart et al. 2020) and a large cluster randomized trial (CRT) was implemented between February 2018 and September 2019 (Betancourt et al. 2020; Jensen et al. 2021). Pre- to post-intervention findings demonstrated that *Sugira Muryango* led to improvements in caregiver behaviors linked to child development and health as well as reductions in violence, which were sustained 12 months after the intervention, at which time improvements in child development were observed.

The Research Program on Children and Adversity in the Boston College School of Social Work is led by Dr. Theresa S. Betancourt and will, in partnership with the University of Rwanda, FXB-Rwanda and Laterite, conduct a longitudinal follow-up study to investigate the longer-term outcomes of the *Sugira Muryango* intervention in families who participated in the CRT. The four-year follow-up will examine the long-term and sustained outcomes of the intervention. In particular, we will look at key indicators of long-term positive outcomes for children such as school readiness and transition to formal schooling. Given the lack of longitudinal research on intervention programs supporting ECD in sub-Saharan Africa, this study will contribute greatly to the body of knowledge on the costs and benefits of investments in ECD and guide policy makers and government leaders on making impactful investments in children, leading to long-term benefits for the population at large.

The follow–up study involves two activities:

> Activity A: Pilot to assess measures performance of newly added measures and field test study protocols.
>
> Activity B: Four-year follow-up of families who participated in the CRT of the *Sugira Muryango* intervention.

2. Study Aims and Objectives

The aim of the study is to assess the impacts of Sugira Muryango on treated families compared to families receiving only usual care (controls) four years after the intervention. Moreover, at this visit, we will also for the first time examine potential spillover effects of Sugira Muryango onto siblings of children enrolled in the original study. The longitudinal follow-up study will assess whether and to what extent Sugira Muryango has an impact on caregivers' awareness and utilization of available services; support for children's education and playful learning; home hygiene; parenting practices, including child feeding/nutrition, stimulation and sensitive care; experiences of intimate partner violence; use of harsh discipline; mental health, including depression and alcohol use; gender attitudes; and overall wellbeing. It will also assess whether and to what extent Sugira Muryango has an impact on children's physical development; cognitive and linguistic development; temperament; enrollment in early or formal education; school readiness, including self-regulation and early literacy and numeracy; mental health, including internalizing and externalizing and depression; behavior, including conduct problems; and gender attitudes. Finally, the study will assess potential covariates such as the impact of the COVID-19 pandemic and associated lockdowns, child disability, and household size and composition.

The specific objectives are:

1. To assess the long-term (4 years post-intervention) impact of Sugira Muryango on caregiver behaviors, attitudes, mental health and wellbeing, especially as related to parenting and intimate partner violence.
2. To assess the long-term (4 years post-intervention) impact of Sugira Muryango on previously enrolled children's physical and cognitive development and health.
3. To assess the long-term (4 years post-intervention) impact of Sugira Muryango on previously enrolled children's behavior, mental health, school readiness and gender attitudes.
4. To assess the long-term (4 years post-intervention) impact of Sugira Muryango on younger and older siblings of children enrolled in the intervention, including on their physical and cognitive development and health, and in the case of older siblings their school readiness, behavior and gender attitudes.

The following activities will be conducted during the proposed 4-year follow up study:

- Activity A: Pilot to assess measures performance and field test study protocols.
    - Phase A1: Translation and adaptation of newly selected measures.
    - Phase A2: A Pilot study of selected new child measures with 150 children and their primary caregivers in a sector of the *Sugira Muryango* PLAY Collaborative Expansion Study to test the feasibility and validity of new tools.

- Activity B: Four-year follow-up of families who participated in the Cluster Randomized Trial (CRT) of the *Sugira Muryango* intervention.
    - Phase B1: Household tracking and re-enrollment of 1,049 households that participated in the prior CRT (April-May 2022).
    - Phase B2: Quantitative (full sample) and qualitative (subsample) data collection with the 1,049 households from the prior CRT, including children and caregivers who were in the original CRT, and siblings in a subset of households to test for spillover effects (June [Quantitative] & December [Qualitative] 2022).

Our **specific hypotheses** are as follows:

Longitudinal Hypotheses

1) *Sugira Muryango* will have effects (superior when compared to usual care households) on a range of outcome domains, both new domains and those assessed in the prior CRT.
    a) *Sugira Muryango* will continue to have effects among eligible caregivers and children compared to controls on a range of outcomes assessed at previous waves of data collection, and
    b) *Sugira Muryango* will have effects on new outcomes that have become relevant as the children have aged, including aspects of school readiness.

Spillover Hypothesis

2) *Sugira Muryango* will have positive effects on younger and older siblings of children who were eligible for and participated in the intervention compared to siblings in usual care households.

Secondary exploratory analyses

Analyses evaluating potential **child and caregiver sex differences** will be performed across all outcomes to examine differences in parenting behaviors and child outcomes as children get older and approach school age.

Study benefits and justification

Potential societal benefits of the study include increased knowledge about evaluating ECD and parenting in Rwanda, increased knowledge of ECD in Rwanda, and increased knowledge regarding family-based ECD interventions to improve child development outcomes in low-resource settings. Intervening in early childhood has been demonstrated to be highly cost-effective for improving child development and life outcomes, yet interventions in low-resource settings—particularly in sub-Saharan Africa—are limited and not always well-evaluated or systematically implemented (Britto et al. 2017). At the individual level, families who participate in the study will

receive free disability and behavioral screening for eligible children and will be connected with support services should a disability or other risk of harm be identified.

The key research question is whether Sugira Muryango provides lasting benefits to children in households that received the intervention. Little longitudinal research into home-visiting ECD interventions in this region exists; findings from the proposed research will add critical evidence to inform Rwanda's expansion of ECD support to families as well as learnings for other countries.

The study has been reviewed and approved by the Rwanda National Ethics Committee, the National Institute of Statistics of Rwanda, and the National Council for Science and Technology in Rwanda.

3. Methodology

Sampling

Target population

The target study population (referred to as "beneficiaries") are families who participated in the Sugira Muryango CRT in 2018-19 (Trial Registration Number NCT02510313). Family inclusion criteria for the CRT were (a) living in the Rubavu, Ngoma or Nyanza District of Rwanda, (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 was willing to discuss and enhance their caregiving practices by interacting with a home-visiting coach (a community-based volunteer or CBV). Further caregiver inclusion criteria were: (a) was aged 18 or older and cared for child(ren) and (b) lived in the same household as the child(ren). Single and dual caregiver families were enrolled to reflect population dynamics and other legal guardians, including parents, aunts, uncles, grandparents, and/or foster parents, were enrolled if no biological parent lived in the household. To ensure that families enrolled in the CRT were socio-economically vulnerable and VUP-eligible, they were randomly selected from an administrative records list, part of the government social protection information infrastructure.

Other than not meeting inclusion criteria, specific family exclusion criteria for the CRT were a) caregiver(s) having severe cognitive impairments which precluded their ability to speak to the research questions under study, b) families or caregivers being in the midst of crisis (e.g., a caregiver(s) with active suicidal attempts or psychosis). Families with ongoing crises or disabilities were excluded from the study and were referred to appropriate services.

For the current **Longitudinal Study, inclusion criteria for families** are (a) having participated in the prior CRT, (b) living in Ngoma, Nyanza, or Rubavu districts, (c) having at least one child who participated in the CRT and who is currently living in the household. **Exclusion criteria**

are: (a) caregiver having severe cognitive impairments which precluded their ability to speak, to understand the research questions, or to get involved in program activities, or (b) families or caregivers who are in the midst of crisis (e.g., a caregiver with active suicidal attempts or psychosis).

Due to the passage of time between waves of data collection, we expect some households to have changed in composition between the one-year post-intervention wave and the current (longitudinal, four-years post-intervention wave). In such cases the following replacement criteria will be observed:

- The current primary caregiver (the person who spends the most time with the child) responds to the caregiver report on self, caregiver report on household, and caregiver report on child. If the current primary caregiver did not participate in the CRT, this will be flagged in the dataset. Field preparation for the study found 72 households where this is the case.
- The secondary caregiver who participated in the CRT responds to the caregiver report on self and caregiver report on household if they are available and eligible (still living in the household), even if they are no longer acting as the secondary caregiver. If the secondary caregiver who participated in the CRT is not available, the current secondary caregiver (if one exists) responds to the caregiver report on self and caregiver report on household.

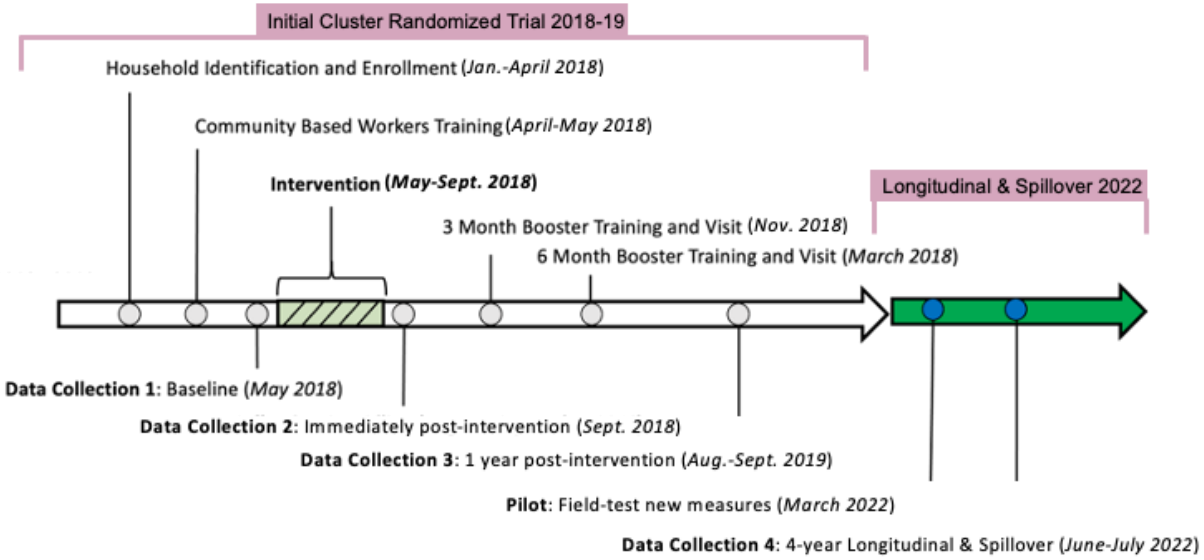
In order to test the Spillover hypothesis, the current study will also enroll siblings of CRT-eligible children in a subsample of eligible households. The sampling methodology is described below. Inclusion criteria for younger siblings are: (a) aged 3 months or older; (b) was not eligible to participate in the original CRT due to age (below 6 months or not yet born at the time of CRT). Inclusion criteria for older siblings are: (a) aged 12 years or younger; (b) lived in the household but was not eligible to participate in the original CRT due to age (above 36 months at the time of CRT); (c) currently lives in the household that participated in the CRT. Although caregivers and children who participated in the CRT will be eligible for the follow-up study regardless of caregiver's presence in the home, siblings will be excluded from sampling for the spillover study if the primary caregiver has been away from the household for six months or more.

The reference period

The CRT began in February 2018 with enrollment, baseline data collection, and subsequent implementation of the Sugira Muryango intervention with treatment families. Immediate post-intervention data were collected in August-September 2018. Families received a three-month post-intervention booster session in November-December 2018 and a six-month post-intervention booster session in March 2019. In addition, a follow-up assessment was completed 1 year post-intervention in August-September 2019.

The current longitudinal study will conduct a new follow-up assessment approximately 4 years after the intervention ended. The pilot tool-testing activity enrolled families in the usual care group (UC) of the Sugira Muryango PLAY Collaborative Study (Trial Registration Number NCT04257383) (these families did not participate in the original CRT). Piloting took place in March 2022. Full quantitative data collection will take place with the longitudinal sample in June 2022.

Figure 1: Sugira Muryango study timeline



Geographical coverage

The original CRT took place in Ngoma, Nyanza, and Rubavu Districts of Rwanda. The same families will be followed up in the proposed longitudinal and spillover study, in the same geographies. Families in which the entire family has moved out of the district they lived in during the CRT will be excluded.

Figure 2: District map

Sample size determination and sampling

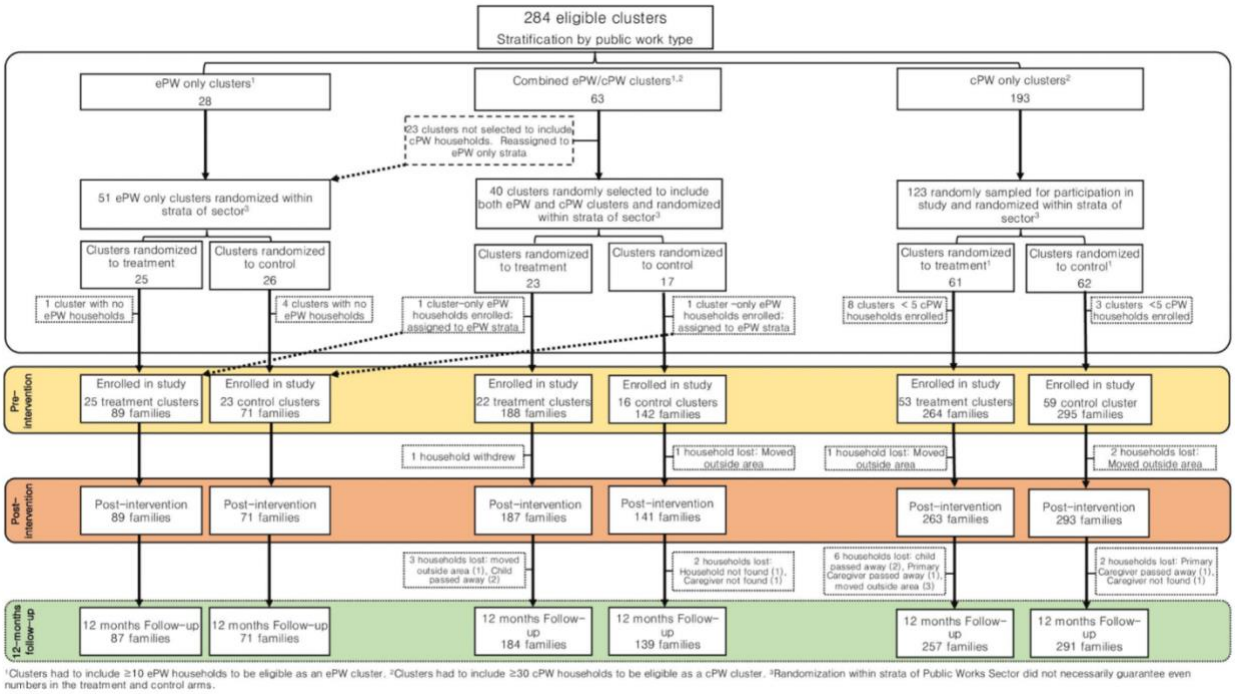
Longitudinal Sample (follow-up with CRT sample)

Given that this is a follow-up study, all 1,028 families who participated in the one year post-intervention wave of the CRT, who meet current inclusion criteria, will be considered the sample of interest. The following values were used to calculate statistical power when sampling for the CRT. 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. Details on the original sampling strategy for the CRT can be found in Betancourt et. al. 2020, and Jensen et. al 2021, and are also summarized below.

Families for the CRT were randomly selected from government-provided lists of families who were eligible for the Vision Umurenge Program (VUP). Families' participation in the VUP and selection into one of two versions of the VUP program, namely classic public works (cPW) or expanded public works (ePW), was determined by governmental policies and was not under the control of the research team. Non-overlapping, geographically defined clusters were created from one or more contiguous villages so one community-based volunteer (CBV) could provide services to all participating families in the cluster. Each cluster comprised at least 30 families participating in the classic Public Works (cPW) program or ten families participating in the expanded Public Works (ePW) program, with some clusters containing both ≥30 cPW and ≥10 ePW households. Randomization to either treatment or control conditions occurred at the cluster level within strata defined by public works type (ePW only, combined ePW/cPW, and cPW only)

and geographic sector. After cluster randomization, 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 (see Cluster Sampling Strategy, Figure 3). The final CRT sample comprised 1,049 households, with 48.41% (n=508) in the usual care (UC) and 51.57% (n=541) in the treatment (*Sugira Muryango*, SM) group. Within households, we sampled n=1084 children ages 6-36 months and n=1498 primary and secondary caregivers. The longitudinal study will sample these same families from the CRT.

Figure 3: Cluster Sampling Strategy and Flowchart of Participants in the Sugira Muryango Cluster Randomized Trial



Spillover study - sibling sample

As discussed above, this study also incorporates a Spillover study examining the effects of the intervention on younger and older siblings of currently eligible children who were not eligible to be enrolled when the original intervention was implemented. Since Sugira Muryango is meant to work with the whole family, and teach beneficiaries transferable skills, we expect siblings of previously enrolled children to also benefit from the intervention. The sibling spillover study will allow us to test hypothesis 2: "*Sugira Muryango* will have positive effects on younger and older siblings of children who were eligible for and participated in the intervention compared to siblings in usual care households".

To select children for the spillover study we will sample younger (age 3 months and older) or older (up to age 12) siblings of the children that were eligible for the original CRT (for qualitative interviews, siblings up to 15 years old may be sampled). For statistical power efficiency, and to prevent a second level of clustering at the household level, younger and older siblings will be randomly selected from different households (only one sibling per household). We will seek to enroll 50% male and 50% female siblings across age categories outlined below. We will seek to balance gender within each village and cluster.

Sampling plan for siblings:

		3 months -2 years	2-4 years	8-10 years	10-12 years
Intervention	Female	31	31	31	31
	Male	31	31	31	31
Control	Female	31	31	31	31
	Male	31	31	31	31

The main outcome of interest for siblings is the child development score, a continuous outcome. We are interested in testing whether the means in the original control and treatment groups are different. The appropriate statistical test for this outcome is a two-sample equality test. The null and alternative hypotheses to be tested are as follows, with $\mu_{c,t}$ the mean in the control and treatment groups, respectively:

H_0: The difference in means of the two groups is equal to 0, $\mu_t - \mu_c = 0$

H_1: The difference in means of the two groups is different from 0, $\mu_t - \mu_c \neq 0$.

Table 1: Parameter for sample size calculation

Parameter	**Value**
ICC (rho)	$\rho = 0.01$
Significance level	$\alpha = 0.05$

Statistical Power	$1-\beta = 0.8$
Standardized Minimum Detectable Effect	$\frac{\mu_t-\mu_c}{\sigma} = 0.25$, where σ is the standard deviation of the outcome, μ_c is the mean in the control group and μ_t is the mean in the treatment group

The relevant parameters needed to calculate the appropriate sample size to conduct this test are outlined in Table 1. For the level of significance and statistical power, the social sciences standard values of α=0.05 and power=80% are used. The standardized minimum detectable effect (MDE) is estimated based on a target of 0.3 standard deviations maximum or smaller, as interventions of the type of Sugira Muryango have achieved effects of this magnitude in the past with similar outcomes. To ensure adequate power, the standardized MDE for this study is set slightly lower at 0.25 standard deviations.

The following formula for a two-sample two-sided hypothesis test is used to compute sample size in each group (control and treatment) based on above parameters (Chow et al. 2008):

$$n_{c,t} = \left(\sigma \frac{z_{1-\alpha/2} + z_{1-\beta}}{\mu_t - \mu_c}\right)^2 \times (1 + \frac{1}{\kappa})$$

Where,

- n is sample size
- $\kappa = \frac{n_c}{n_t}$ is the matching ratio and is equal to 1
- $\frac{\mu_t-\mu_c}{\sigma}$ is the standardized MDE
- α is Type I error
- β is Type II error, meaning $1-\beta$ is power
- $z_{1-\frac{\alpha}{2}}$ is the critical z value for $1-\frac{\alpha}{2}$
- $z_{1-\beta}$ is the critical z value for $1-\beta$

$$n_{c,t} = \left(\frac{1}{0.25}(1.96 + 0.84)\right)^2 \times 2 = 250$$

With those parameters, we find a total sample size of 500 siblings, with 250 siblings in control clusters and 250 siblings in treatment clusters.

As the sampling strategy includes clustering, the sample size calculation also takes into account the effects of the within and between cluster variation in the calculation of the required effect

size. The formula below is used for the calculation of the design effect where m is the size of each cluster and ρ is the Intra-Cluster Correlation (ICC) (Collett 2014):

$$DEFF = (1 + (m - 1)\rho)$$

The average number of households per cluster is 5. The sibling sample will be drawn from all available clusters in the study to maximize power. As this study sample includes 197 clusters, we can calculate: $m = \frac{500}{197} = 2.54$ participants per cluster. In practice this means that some clusters will include 3 siblings while others will include 2 siblings.

The ICC estimate is taken from the initial CRT and is estimated to be 0.01 at the cell level. Such a small ICC implies that the clustered design does not result in significant data heterogeneity between clusters. Note that actual clusters for this study are between village and cell size, and do not directly align with Rwanda's administrative units. However, for the purpose of power calculations, we assume the ICC for the cell is comparable to the ICC for the actual cluster size.

Using this, we calculate a design effect as follows:

$$DEFF = (1 + (2.54 - 1)0.01)$$

$$DEFF = 1.0154$$

Due to the very small ICC, the effect of the design on the sample size is very limited. Given that we use for this calculation a MDE of 0.25, versus the target of 0.3, the calculation will provide a conservative (slight overestimate) of the sample needed, and so we can assume the calculated sample size accounts for this very limited design effect.

Therefore, with 197 clusters (k), we will sample 2 to 3 siblings (m) per cluster for a total sample size of 500 siblings (k*m=n), with 250 siblings in control clusters and 250 siblings in treatment clusters.

Qualitative sampling

From the quantitative sample, 90 caregivers (30 from single-caregiver homes [mostly female; if we find single-male-headed homes in our sample, some of these will be purposively included in the sample], 30 male caregivers from dual-caregiver homes, and 30 female caregivers from dual-caregiver homes) will be selected to additionally participate in qualitative interviews. Families will be selected as high- and low-responsive to the intervention based on quantitative data. We will also sample 60 CRT-eligible children and/or older siblings from the same households. Siblings will be selected to participate in qualitative interviews (60 children in total). Selection of children will be stratified by gender (50% male and 50% female).

Analysis Plan

Outcomes: CRT-eligible children (longitudinal sample)

Continuous Outcomes

The goal is to determine the efficacy of SM four years from study initiation. Measurements of most outcome variables (see section 6) were taken at baseline, immediately post-intervention, and 12 months post-intervention. In this study, several of these measures will be collected again four years post-intervention. Subjects in the CRT were cluster randomized (Section 4, "The sample size and its calculations and the sampling methods used") into a treatment group (SM) and usual care (UC). Continuous outcomes collected during the original CRT will be analyzed using linear mixed-effect models and intervention effects will be evaluated based on the differences in slopes between SM and UC. The linear mixed-effect model, with a continuous outcome variable (e.g., KABC-2 total score) as the response variable, will include the treatment status (SM vs UC), time, and their two-way interaction as fixed effects. Mixed effect models, also known as hierarchical linear models (HLM) or multilevel 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 clusters (i.e., geographical regions). 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, although these changes were neither expected nor hypothesized, a quadratic term for time will be added as a fixed effect in order to explore possible changes (acceleration) in the rate of growth. Lastly, even though in prior analysis no differences between VUP programs were found,the type of VUP program will be modeled as a fixed effect to evaluate potential long-term impacts of the VUP program. A mathematical representation of this model is:

$$Y_{ijk} = B0 + b0_i + B_1*\text{TreatmentGroup} + B_2*\text{Time} + b_{1i}*\text{Time} + B_3*\text{Time}*\text{Treatment group} + b_{2k} + b_{3k}*\text{time} + B4_m*\text{PubicWorksGroup} + e$$

Where, Y_{ijkm} represents the value of the i^{th} subject (i = 1 . . . I), at the j^{th} time point (j=1, . . . 4), in the k^{th} cluster (k = 1, . . 200) and m is the VUP indicator with levels expanded public works, classic public works and mixed. The set of $b0_i$ variables represent subject-specific intercepts and are assumed $b0_i \sim N(0, \sigma I)$. The b_{1i} 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 b_{3k} are region-specific slopes and assumed $b_{3k} \sim N(0, \sigma s_{region})$. The value e represents residual variation and is assumed $e \sim N(0,\sigma)$. Model efficacy will be evaluated using the likelihood ratio test for inclusion of the Time*Treatment interaction (null hypothesis $B_3 = 0$). Random effects will be tested during the model building but we will seek to simplify models where appropriate. For example, we will 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 full maximum likelihood. Because of randomization, we have controlled in theory for imbalances in observed and unobserved variables (confounders). Moreover, previous analyses from the CRT sample have shown that group differences at baseline tend to be minimal. However, maturation and developmental processes might play a role so models with some key demographic variables will be evaluated and models will be compared with the main results.

Model assumptions and alternative tactics: Model assumptions will be tested. 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 subjects. In the case of non-normality of outcomes, we will consider response transformations or the use of models that do not make distributional assumptions (e.g., generalized linear mixed effect models). If the linearity assumption does not hold across time (baseline, immediately post-intervention, one year post-intervention, and four years post-intervention) marginal effects of the treatment at each timepoint will be estimated, focusing on the 4-year group differences adjusted for baseline. Lastly, if the examination of panel plots reveals a large amount of heterogeneity among trajectories of the outcomes by subjects, Growth Mixture Models will be used in order to 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. Model 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 it is found that outliers are having an oversized effect on the estimated coefficient values, preference will be given to the models where the outliers are removed. However, both sets of results will be presented.

Binary Outcomes

The analysis for these outcomes will be similar to the analyses above with accommodations being made for the binomial distribution of the outcome. In particular, mixed-effect logistic regression models for binomial distributions and a logistic link function will be used, a specific form of generalized linear mixed models (GLMM). Outside of accounting for the correct distribution of the response variable, the model nesting structure and modeling strategy (i.e., covariates, fixed, and random effects) will be the same as the one outlined in the previous section for the continuous outcomes.

Model Assumptions and Alternative Tactics: We will examine Pearson residuals to assess the model 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, a Wald-type type test for regression coefficients will be considered. 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

Prior CRT data examination and analysis of outcomes like harsh disciplinary actions or physical and sexual abuse as suffered or inflicted showed that these variables were not normally distributed but resembled a Poisson distribution, with a great concentration of zero values. In fact, previous 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 2006; 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 that 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.

Linking procedures to address CRT sample children's maturation

The current study design addresses developmental changes and maturation effects by using age-appropriate instruments to measure the constructs of interest. This posits a measurement challenge given some of the children that during the CRT were assessed using a certain instrument (e.g., ASQ-3) will now need to be assessed using a different age-appropriate tool (e.g., KABC-II) considering they are currently out of the original instrument recommended age span. To address this, statistical methods for test linking will be used. Linking is an umbrella term covering different processes and techniques that involve transformations between the scores

from one test and those of another in order to achieve comparability across different instruments. In general terms, these transformations can then be divided into three big categories: Predicting, Scale Aligning, and Equating (Holland 2007). In particular, considering 1) instruments are measuring the same construct 2) scores come from a common population (single-group design), and 3) reliability and difficulties may vary between test forms, equipercentile equating techniques will be used. In short, as described by Gonzalez and Wiberg (2017), with equipercentile equating, test form X is equated to test form Y by identifying X scores on test form X that have the same percentile ranks as Y scores on test form Y. This will allow for the transformation of scores from two different tests onto a common scale. These scores will be later analyzed using the three-level linear mixed model for continuous outcomes as specified before.

Analysis of new outcomes (CRT and Siblings samples)

Outcomes that were not originally collected from eligible children during the original CRT (e.g., school readiness) and outcomes for younger and older siblings who were not assessed during the original CRT will be analyzed using either linear mixed models (continuous outcomes), mixed-effect logistic regression models (binary outcomes), or mixed-effect Zero Inflated Poisson (ZIP) models (count outcomes). These models will have two levels of nesting: individuals nested within randomization clusters. Considering these subsamples will have data from only one measurement wave, the longitudinal nature of previous models (i.e., individuals nested within measurement waves) is not present. In this particular case, intervention effects will be evaluated using the coefficient for the treatment group variable (SM vs UC). As before, since we expect region-level effects, subject-specific slopes and intercepts will be modeled as random effects nested within randomization cluster, also modeled as a random effect. And the type of VUP program will be modeled as a fixed effect to evaluate potential long-term impacts of the VUP program. Although the use of one timepoint is not ideal to evaluate intervention effectiveness, this study will take advantage of the original randomized design (see Section 4, Subsection "Sample Selection"), which generated comparable intervention groups that are supposed to be similar in all relevant observed and unobserved characteristics except for the treatment allocation. In short, randomization ensures no a-priori knowledge of treatment allocation, and also that groups do not differ in any systematic way in terms of observed or unobserved characteristics (Altman and Bland, 1999; Suresh 2011). Lastly, in order to account for potential imbalances between the siblings' treatment and control subsamples or imbalances due to changes in household structure and other potential observed confounders, group equivalence analysis will be performed. When statistically significant imbalances are found those variables will be included as statistical controls in the respective mixed model fitted.

Missing Data

Data will be analyzed under the Intention-to-Treat approach, a data analysis strategy according to which all participants are included in the analytic sample and considered part of the group to which they were originally randomized, regardless of intervention completion. This will help in

preventing bias due to sample loss. Missing data, either from attrition or from survey non-response, will be addressed using HOTDECK multiple imputation techniques. Lost cases will be replaced by randomly selecting cases matched on key demographic, and socioeconomic variables such as age, gender, educational attainment, household structure, and also the treatment group.

4. Measures

Table 2: Caregiver measures

Caregivers				
Data collection type	**Outcome**	**Measure**	**New measure**	**Outcome type**
Caregiver report on household (taken at home)	Household composition	"Household composition" survey	No	Covariate
Caregiver report on self (taken at home)	Service utilization	"Service utilization" survey	Yes	Secondary
Caregiver report on household	Support for child's education	Epstein survey	No	Secondary
Caregiver report on household (taken at home)	Financial decision making; father engagement	DHS/Promundo "Father Engagement & Financial Decision-Making" survey	No, but some changes	Secondary
Caregiver report on household (taken at home)	Stimulating care/play	"The Role of Play in Children's Learning" survey	Yes	Primary
		"Parental Acceptance-Rejection Questionnaire"	No	Secondary
Caregiver report on household (taken at home)	Home sanitation & hygiene	"Multiple Indicator Cluster Survey (5)" WASH module	No	Secondary
Caregiver report on household (taken at home)	COVID-19 pandemic impact	"COVID Questionnaire"	Yes	Covariate
Caregiver report on self (taken at home)	Caregiver conflict; victimization &	DHS "Intimate Partner Violence" survey	No	Primary

	perpetration of violence			
Caregiver report on self (taken at home)	Depression/ anxiety	"Hopkins Symptom Checklist (25)"	No	Secondary
Caregiver report on self (taken at home)	Caregiver alcohol problems	World Health Organization "Alcohol Use Disorders Identification Test (AUDIT)"	No	Secondary
Caregiver report on self (taken at home)	Quality of life	"Flourishing Scale"	Yes	Secondary
Caregiver report on self (taken at home)	Views of gender/ masculinity	"Gender Equitable Men" scale "Gender Equitable Men" norms scale	Yes	Secondary
Caregiver report on household (taken at home)	Attitudes towards children's education	"International Development & Early Learning Assessment" parent attitudes module	Yes	Secondary
Caregiver report on household (taken at home)	Expectations towards children's education	"International Development & Early Learning Assessment" parent expectations module	Yes	Secondary
Caregiver report on self (taken at home or central location)	Various	Qualitative interviews	Yes	NA (qualitative)

Table 3: Child measures

Children							
	Outcome	**Measure**	**Younger siblings**	**CRT-eligible children**	**Older siblings**	**New measure**	**Outcome type**

Data collection type							
Caregiver report on child (central location)	Physical development (stunting, wasting, underweight)	Anthropometric measures	Younger	CRT-eligible		No	Primary
Caregiver report on child (central location)	Child development	"Ages and Stages Questionnaire-3"	Younger			No	Primary
Caregiver report on child (central location)	Playful parenting; childcare arrangement	"International Development & Early Learning Assessment" stimulating care and play module	Younger	CRT-eligible	Older	No	Primary
Caregiver report on child (central location)	Child discipline	"Multiple Indicator Cluster Survey" child discipline module	Younger	CRT-eligible	Older	No	Primary
Caregiver report on child (central location)	Feeding practices	World Health Organization "Infant and Young Child Feeding Practices" survey	Younger			No	Secondary
Caregiver report on child (central location)	Enrollment in early education	"Enrollment in early education" survey	Younger	CRT-eligible	Older	Yes	Secondary
Caregiver report on child (central location)	Temperament	"Infant Behavior Questionnaire" ; "Early Childhood Behavior Questionnaire"	Younger			Yes	Secondary

Caregiver report on child	Disability	"Multiple Indicator Cluster Survey" child functioning module	Younger		Older	Yes	Covariate
Observation (at home)	Stimulation	"Home Observation Measurement of the Environment" infants/toddlers survey	Younger			No	Primary
		"Home Observation Measurement of the Environment" early childhood survey		CRT-eligible		Yes	Primary
Observation (at home or central location)	Sensitive care	"Observation of Mother-Child Interactions"	Younger			No	Primary
Observation (central location, in the presence of caregiver)	Developmental milestones & language	"Malawi Developmental Assessment Tool" (under consideration)	Younger			No	Primary
Observation (central location, in the presence of caregiver)	Cognition; language	"Mullen Scale of Early Learning"	Younger			Yes	Primary
		"Kaufman Assessment Battery for Children-2"		CRT-eligible	Older	Yes	Primary
		"Weschler Preschool & Primary Scale of Intelligence"		CRT-eligible		Yes	Primary
Caregiver report on child (central location)	Enrollment in formal education	"Child Enrollment in Formal Education" survey		CRT-eligible	Older	Yes	Secondary

Caregiver report on child (central location)	Internalizing/ externalizing	"Child Behavior Checklist"		CRT-eligible	Older	Yes	Primary
Child report on self (central location, in the presence of caregiver)		Youth Self Report			Older	Yes	Primary
Observation (central location, in the presence of caregiver)	Self-regulation	"Preschool Self-Regulation Assessment"		CRT-eligible		Yes	Secondary
Child report on self	Child discipline	Adapted "Multiple Indicator Cluster Survey" and "ISPCAN Child Abuse Screening Tool"			Older	Yes	Primary
Caregiver report on child (central location)	Child disability	NCDA "Disability Screener"		CRT-eligible		Yes (replaced previous disability screener)	Covariate
Observation (central location, in the presence of caregiver)	Early reading	"Early Grade Reading Assessment"			Older	Yes	Secondary
Observation (central location, in the presence of caregiver)	Early math	"Early Grade Mathematics Assessment"			Older	Yes	Secondary
Child report on self (central location, in the presence of caregiver)	Depression	"Center for Epidemiological Studies Depression Scale"			Older	Yes	Primary

Child report on self (central location, in the presence of caregiver)	Conduct problems	"Youth Conduct Problems Scale - Rwanda"			Older	Yes	Primary
Child report on self (central location, in the presence of a caregiver)	Views of gender/ masculinity	"Gender Equitable Men" scale		CRT-eligible		Yes	Secondary
		"Gender Equitable Men" scale			Older	Yes	Secondary
Child report on self (central location, in the presence of a caregiver)	Various	Qualitative interviews		CRT-eligible	Older	Yes	NA (qualitative)

5. Safety & Adverse Events

Risks to Subjects

This study does not involve a new intervention. Potential risks are therefore limited to risks associated with research participation. Potential risks include:

- In some cases, fatigue from the psychosocial assessment administered may occur.
- There is a small risk of loss of confidentiality.

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.

Monitoring

A risk of harm protocol, including flagged items from the assessment battery, provides a structured protocol for monitoring and responding to adverse effects on participants. This protocol has been reviewed and approved by the Rwanda National Ethics Committee and is used in enumerator training and fieldwork.

6. Study Limitations

A primary study limitation is the potential introduction of confounding effects in the three years since the last data collection (and four years since intervention). In particular, the shock of the COVID-19 pandemic is not yet well-understood. In addition, the targeted families are among the most vulnerable in Rwanda and therefore they may have been enrolled in other interventions targeting similar outcomes in the intervening years. Time effects will be modeled as random effects in the multilevel model, and a quadratic term has been added to the model as a fixed effect to explore possible changes in the rate of growth. Further, questionnaires on COVID-19-related experiences and participation in other programs have been added to the battery to improve understanding of possible covariates related to the pandemic and to other interventions.

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) through two pilot studies (Betancourt et al. 2018; Barnhart et al. 2020). New measures added for this wave of data collection were piloted with demographically similar households in Rwanda, and this pilot indicated strong reliability and validity of the new measures.

7. Compensation, Consent & Assent

Households will receive 5000 Rwandan Francs for quantitative data collection . Households selected to participate in qualitative interviews will receive an additional 3000 Rwandan Francs. All caregivers gave written informed consent for themselves and their eligible children ages 6–36 months to participate in the CRT and will provide consent again for the follow-up study. Children ages four and older will also provide assent for themselves.

8. Provisions for vulnerable subjects

All households in the study had an Ubudehe 1 poverty categorization at the time of intervention. As such, all study procedures are created and implemented with provisions for vulnerable participants. As mentioned above, 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.

9. 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.

10. References

Altman, D. G., & Bland, J. M. (1999). Statistics notes: Treatment allocation in controlled trials: why randomise? BMJ, 318(7192), 1209–1209. https://doi.org/10.1136/bmj.318.7192.1209

Barnhart, D. A., Farrar, J., Murray, S. M., Brennan, R. T., Antonaccio, C., Sezibera, V., Ingabire, C., Godfroid, K., Bazubagira, S., Uwimana, O., Kamurase, A., Yousafzai, A., Betancourt, T. (2020). Lay-worker delivered home promotes early childhood development and reduces violence in Rwanda; a randomized pilot. *Journal of Child and Family Studies*; 29, 1804-1817. https://doi.org/10.1007/s10826-020-01709-1

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*; 190(2), 219-235.

Betancourt, T.S., Jensen, S.K.G., Barnhart, D.A., Murray, S.M., Yousafzai, A.K., Farrar, J, Godfroid, K., Bazubagire, S.M., Rawlings, L.B., Wilson, B., Sezibera, V., & Kamurase, A. (2020). Promoting parent-child relationships and preventing violence via home-visiting: a pre-post cluster randomised trial among Rwandan families linked to social protection programmes. *BMC Public Health*; *621(20)*, 2-11.

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. (2017). Nurturing care: promoting early childhood development. *The Lancet; 389*(10064), 91-102.

Chow, S. C., Shao, J., Wang, H., & Lokhnygina, Y. (2017). Sample size calculations in clinical research. chapman and hall/CRC.

Collett, D. (2014). Modelling Binary Data (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16654

Everitt, B. & Hothorn, T. (2006). A handbook of statistical analyses using R. Chapman & Hall/CRC.

González, V. & Wiberg, M. (2017). Applying test equating methods: using R. Springer.

Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), Linking and aligning scores and scales (pp. 5–30). Springer Science + Business Media. https://doi.org/10.1007/978-0-387-49771-6_2

Jensen, S. K., Placencio-Castro, M., Murray, S. M., Brennan, R. T., Goshev, S., Farrar, J., ... & Betancourt, T. S. (2021). Effect of a home-visiting parenting program to promote early childhood development and prevent violence: a cluster-randomized trial in Rwanda. BMJ Global Health, 6(1), e003508.

Long. S. (1997). Regression models for categorical and limited dependent variables. Sage Publications.

Suresh, K. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. Journal of Human Reproductive Sciences, 4(1), 8. https://doi.org/10.4103/0974-1208.82352

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