

PRE-ANALYSIS PLAN

Growing Stronger Together: A Family Strengthening Approach to Reduce Child Engagement with Armed Forces

April 4, 2026

This pre-analysis plan is meant to accompany the American Economic Association (AEA) Trial Registration #AEARCTR-0016067: *Growing Stronger Together: A Family Strengthening Approach to Reduce Child Engagement with Armed Forces*. The trial was registered in the AEA Trial registry on May 22, 2025, prior to baseline data collection and cross registered on the Pan-African Clinical Trial Registry on June 4, 2025 (PACTR202506718109255). The trial follows a Hybrid Type 1 Effectiveness-Implementation Individual Randomized Controlled Trial (RCT) approach. This design experimentally tests effectiveness, while simultaneously gathering data on implementation factors such as cost, sustainability and acceptability, among others.

This analysis plan lays out the basic analysis choices alongside outcome measures which will comprise the main trial results. At the time of writing of this analysis plan, follow-up data collection for the trial was completed – however not yet accessible or analyzed by the researchers responsible for key decisions on the analysis plan.

1. Objectives and theory of change

Growing Strong Together (GST) is an innovative approach that combines curriculum- and group-based effective parenting programming with tailored support for families who have had children engaged in or at risk of recruitment into armed groups. The GST intervention consists of a parenting curriculum delivered over approximately 3 months through 20 bi-weekly sessions, each lasting between 1 and 3 hours delivered by trained facilitators to groups of parents and caregivers.¹ Parents of at-risk adolescents (approximately 13-17 years) are recruited to take part in the intervention, forming groups of approximately 20 caregivers. Intervention facilitators include International Rescue Committee (IRC) child protection staff and trained community volunteers (each group is led by one IRC facilitator supported by one female and one male community volunteers who will receive 8 days of training on GST facilitation). GST session topics include specific topics related to engagement in arms groups and forces (e.g., providing information on who is at risk, motivations for joining, experiences within armed groups, and consequences for children), improving parenting skills specifically for highly vulnerable children (e.g., communication, listening, managing emotions, and positive discipline), and skills to improve caregiver wellbeing (e.g., self-care practices to reduce stress, grapple with stigma, identifying and addressing their own emotions, and identifying sources of social support).

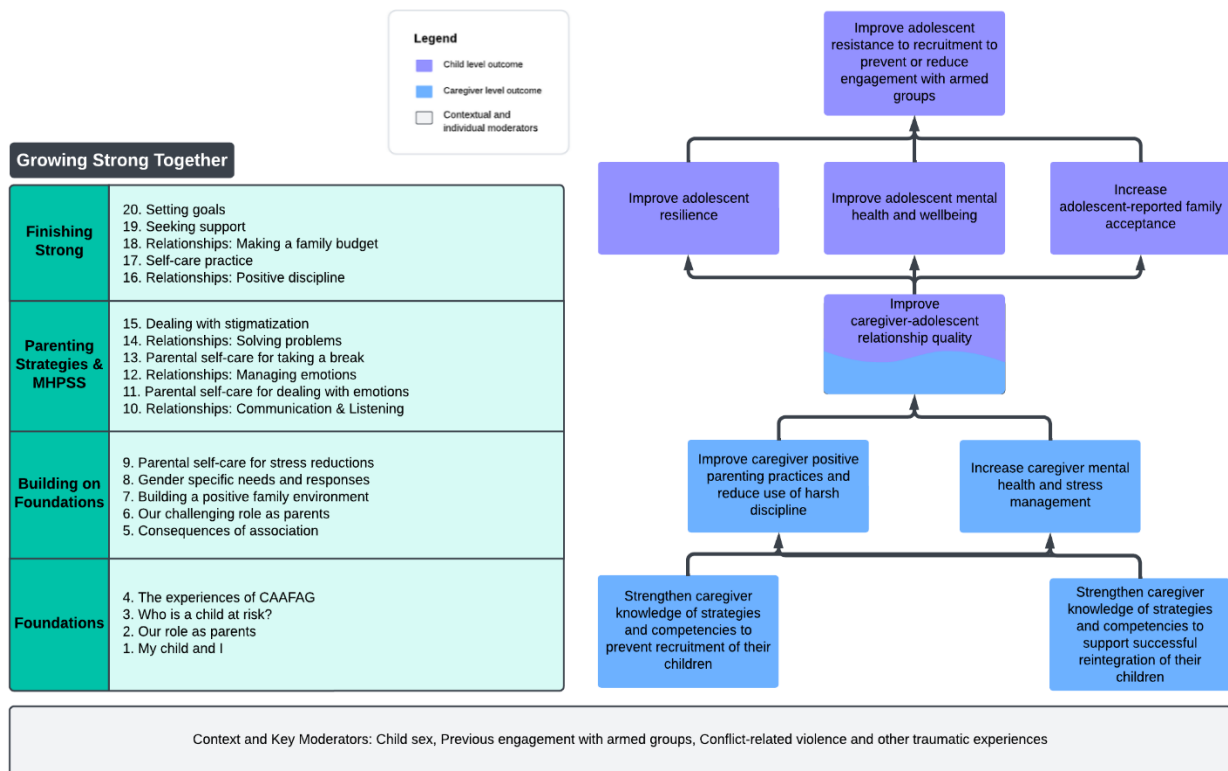
¹ GST also includes period home visits for facilitators to check in on the family and provide any tailored support – however in practice, in the CAR implementation, these occurred only once throughout the intervention period.

The overall objective of the trial is to answer the following questions:

1. What is the effectiveness of GST in preventing child recruitment by armed forces and armed groups and supporting their reintegration into communities?
2. Through which potential mechanisms does GST operate (e.g., promoting parenting skills, improving caregiver-child relationship quality, child resilience, or mental health)?
3. What are the feasibility, acceptability, and implementation factors needed for successful implementation of GST?

The focus of this analysis plan are on questions (1) and (2), while question (3) will be answered primarily with qualitative and monitoring data. The impact analysis follows the GST theory of change, developed for this trial (**Figure 1**). In the theory of change, caregivers who participate in the intervention are expected to gain knowledge of strategies and competencies to prevent recruitment of their children, as well as support successful reintegration (*bottom row, blue boxes*). In turn, these competencies are expected to improve caregiver positive parenting, reduce use of harsh discipline (harsh parenting) and increase caregiver mental health. These positive impacts are hypothesized to improve caregiver and adolescent relationship quality (*middle box, blue and purple*). Relationship improvements lead directly to improved adolescent resilience, mental health, family acceptance, and finally improve their resistance to recruitment and engagement with armed groups (*top row, purple box*). Key moderators of these relationships are: i) child sex, ii) previous engagement with armed groups and iii) exposure to conflict-related violence and other trauma experiences. We describe the measurement of these key indicators further in later sections.

Figure 1. Growing Stronger Together theory of change



2. Methodology

2.1 Sample

The quantitative sample was based on the recruitment of caregiver-child dyads with an estimated total sample of 450 dyads across six villages at baseline (split equally by study group: 225 treatment and 225 control, or delayed entry). This sample size was needed to estimate a Cohen's d of 0.25 on key impact and outcome indices (i.e., resilience to recruitment, adolescent mental health, or family acceptance), considering a predictive power of covariates of 0.4. Intervention drop out and attrition rates were estimated at 30% and 15% respectively, and the usual parameters of 80% power and alpha <0.05 were used.

At baseline, the total matched dyad (caregiver-child) sample was 447 + 11 caregiver only observations, for a total randomized baseline sample of 463. Randomized allocation was 1:1 across study groups using Stata (computer based) stratified by village and adolescent sex. The primary analysis sample for the study consists of both child-level and caregiver-level longitudinal samples, originating from this baseline sample (as discussed further below).

2.2 Analysis

Analysis will be undertaken using the panel sample of caregivers and children who were interviewed at baseline and were re-contacted again at follow-up. Thus the analysis will be carried out on the non-dyad ("unbalanced") sample to maximize sample size, however a robustness check will be carried out on the dyad sample and reported in Supplementary Materials. Standard attrition analysis (overall and differential) will be conducted to ensure the internal validity of the evaluation, and baseline balance of background characteristics and outcome variables will be conducted to demonstrate success of the randomization. If selective attrition is shown to be a problem, we will include robustness checks using lee bounds following Lee (2009).

Impact analysis will follow standard analysis of covariance (ANCOVA) models, with a generalized form as follows:

$$(1) \quad Y_{cv1} = \alpha + \beta_t GST_c + \gamma Y_{cv0} + \delta X_{cv0} + \sigma_v + \varepsilon_{cv}$$

Where Y_{cv1} is the outcome of interest for caregiver (or child) c from village v at follow-up and Y_{cv0} is the outcome of interest at baseline. GST_c equals one if the caregiver(child) is in the GST treatment arm. The β represents the intent-to-treat (ITT) estimator, or the effect of being assigned to GST. We plan to present both unadjusted results (controlling for village and sex, the randomization strata), as well as those controlling for X_{cv0} , a vector of background characteristics of the caregiver(child) measured at baseline.

The background (baseline) characteristics we plan to include in equation (1) are as follows:

- *Caregiver analysis:*
 - Age (years)
 - Sex (most caregivers are female)
 - Number of household members
 - Household wealth index

- Currently displaced (=1)
- *Child analysis:*
 - Age (years)
 - Sex (most children are male)
 - Disability status (moderate or severe=1)
 - Currently in school (=1)
 - Number of household members
 - Household wealth index
 - Currently displaced (=1)

In addition to the above covariates, all regressions will control for village fixed effects as randomization was stratified on village (σ_v). All regressions will cluster standard errors at the village level, however a robustness check will be done with robust standard errors. We will run linear probability models for binary outcomes and ordinary least squares models for continuous variables.

2.3 Robustness checks

We plan to conduct a number of analysis to extend and explore the robustness of impacts.

- *Heterogenous impacts:* We will conduct heterogeity analysis by interacting the treatment indicator (GST_c) in equation (1) with key baseline moderators: i) child sex, ii) child previous (ever) engagement with armed forces and groups and iii) household exposure to conflict-related violence. These analyses should be considered exploratory, as the trial was not powered to detect differential treatment effects.
- *Treatment on the treated (TOT):* As not all caregivers will attend all group sessions, we are interested to know if intensity of treatment matters for impacts. As a robustness check we will explore TOT measures using both “any attendance” and intensity of treatment (i.e. high attendance, if there is sufficient variation in attendance). All other parameters of the estimation will remain the same.
- *Disaggregating impacts on resilience to recruitment:* As the resilience to recruitment scale is one of the key impact outcomes, we plan to analyze disaggregated items as individual standardized z-scores to understand impact dyanmics (and report in Supplementary Materials). Due to the large number of indicators, we will correct for multiple hypothesis testing (MHT) using false discovery rate q-values (Simes, 1986).
- *Social desirability bias:* As many of the outcomes we focus on may be subject to social desirability bias, we will conduct a robustness check to see if individuals reporting more socially desirable tendencies are driving potential treatment effects. Following Dhar and colleagues (2022), we will interact the treatment indicator (GST_c) with the Marloe Crowne score of “social desirability.” The score is comprised of questions measuring the caregiver(child) motivation to conform with socially accepted actions. As this indicator was only collected at follow-up, we will first demonstrate the treatment does not affect the scale itself, prior to running the sensitivity analysis.

Other exploratory analyses may be conducted as part of the main trial paper, however will be clearly indicated as not pre-specified (and rather “exploratory”).

2.3 Missing data

We plan to handle missing data using the following protocol:

- ***Outcome indicators***: For all individual indicators that make up outcome scales (or indices) we will replace missing data with the mean of the remaining study group and report the percentage missing – as long as there is at least one non-missing item in the scale. For non-scale missing outcome variables at baseline, we will similarly replace missing data with the mean of the remaining study group and include an indicator for “missing” at baseline. We will not replace stand-alone (non-scale) missing outcome variables at follow-up.
- ***Baseline background characteristics***: For baseline background indicators that are part of scales (i.e., wealth index), we will replace missings in a similar way to outcome indicators, replacing a single item with the mean of the study group and report the percentage missing -- as long as there is one non-missing indicator that comprises part of the scale. For non-scale baseline background variables, we will explore ways to replace indicators using follow-up data or information from dyads. For all remaining covariates, we will replace missings with the study group mean and include an indicator for “missing” at baseline.

3. Outcome indicators

The key indicators for the trial fall into three groups following the theory of change: 1) primary impacts (child level), 2) secondary impacts (child level), 3) secondary impacts (caregiver level). Outcomes in indices will be equally weighted aggregates and standardized to the control group using z-scores. For all indices, we will first run scale reliability analyses to ensure scales were appropriately adapted for the study population – and may drop single items if they show poor fit.

Table 1 defines child level outcomes (groups 1 and 2) and Table 2 defines the caregiver level outcomes (group 3). Items in italics (*and marked with *asterisks*) are exploratory, as they are either supplementary to the main theory of change, or based on indicators which have not previously been validated. These outcomes will be analyzed as additional exploratory measures to support the main outcomes. All indicators in Table 1 are reported by the focal child, and in Table 2 are reported by the caregiver, unless otherwise noted.

Table 1. Child level primary and secondary impacts

Group	Indicator	Description and construction of the variable
Primary impacts	Any engagement	<ul style="list-style-type: none"> • Binary indicator of reported association with armed forces or groups in the last 3 months (baseline phrasing “currently associated”) (both child and caregivers reported)
	Resistance to recruitment	<ul style="list-style-type: none"> • The Resistance to Recruitment scale is a novel 19-item scale capturing attitudes and beliefs that may protect against or increase vulnerability to recruitment and engagement with armed groups and forces, developed as part of the GST formative work and trial (Falb et al. 2026). Responses to each

		<p>item are: i) disagree, ii) neutral, iii) agree, scored on a 0 – 2 scale. The outcome will be analyzed as a continuous, normalized score aggregating all items.</p>
	<p><i>Support for engagement vignettes*</i></p>	<ul style="list-style-type: none"> • <i>Support for child engagement vignettes are exploratory measures developed by trial team capturing responses to hypothetical scenarios assessing decision-making regarding child recruitment. Three vignettes assess: (1) economic pressure, (2) peer pressure, (3) return after leaving. Response options include: i) strongly forbid from joining, ii) try to discourage but accept decision if insists and iii) encourage to join to support family. Analysis of total score will be explored, as well as individual vignettes (both child and caregivers reported)</i>
<p>Secondary impacts</p>	<p>Resilience</p>	<ul style="list-style-type: none"> • The Brief Resilience Scale (BRS) is a 6-item scale measuring the ability to bounce back or recover from stress (Smith et al. 2008). The BRS is scored by averaging all items after reverse-coding items 2, 4, and 6. Items are rated on a 1–5 Likert scale ('strongly disagree' to 'strongly agree'). A final mean score of 1.0–2.99 indicates low resilience, 3.0–4.30 normal, and 4.31–5.0 high resilience. The outcome will be analyzed as a continuous, normalized score aggregating all items.
	<p>Mental Health</p>	<ul style="list-style-type: none"> • The Patient Health Questionnaire for Adolescents (PHQ-A) is an instrument that assesses anxiety, eating, mood and substance use disorders among adolescents (Johnson et al. 2002). The PHQ-A includes 9 items scored on a 4-point Likert scale, ranging from 0 (not at all) to 3 (almost every day), with a maximum total score of 27. The scoring system for depression categorizes scores between 0-4 as non-depression, 5-9 as mild depression, 10-14 as a moderate degree of depression, 15-19 as moderate to severe, and 20-27 as severe. The outcome will be analyzed as a continuous, normalized score aggregating all items. • <i>Hope*: 6-item, Likert-scale examining children's hope for the future as a positive dimension of child wellbeing and mental health (Snyder, et al., 1997). This six item measure is "based on the premise that children are goal directed and that their goal-related thoughts can be understood according to two components: agency and pathways." These two components, agency (ability to initiate and sustain action towards goals) and pathways (capacity to find a means to carry out goals), are assessed by the measure. Adding pathway and agency scores will provide an overall hope score (i.e., level of hope). Scores of 4-8 indicate no to very low hope, 9-12 indicate slightly hopeful, 13-16 indicate moderately hopeful, and 17-24 indicates highly hopeful. The outcome will be analyzed as a continuous, normalized score aggregating the pathway and agency score.</i>

	Perception of Family Social Acceptance	<ul style="list-style-type: none"> The Perceived Acceptance scale is a 12-item scale with 5-point Likert-based responses ('strongly disagree' to 'strongly agree') (Brock, et al 1998). This scale was designed to assess relationship-specific elements of perceived acceptance. Twelve items assess acceptance from family. The outcome will be analyzed as a continuous, normalized score aggregating all items.
	Caregiver-adolescent relationship quality	<ul style="list-style-type: none"> The Positive Parenting Scale – Adolescent is a 9-item scale scored on a 4-point likert scale ranging from 0 (almost never) to 3 (very frequently), range 0-27. The outcome will be analyzed as a continuous, normalized score aggregating all items (Puffer, et al 2017). This scale was adapted from the Positive Parenting Scale developed by Puffer et al. (2017) for adolescent respondents.
	Harsh Discipline	<ul style="list-style-type: none"> The Child Abuse Screening Tool - Trial version is a 14-item scale developed by the International Society for the Prevention of Child Abuse and Neglect (ISPCAN) (Meinck et al. 2018). The scale is meant to capture child abuse (physical, emotional, sexual, neglect) and witness violence in the past month: physical abuse (6 items), emotional abuse (6 items), and neglect (2 items). Each item scored on a binary scale: 0 (not used) or 1 (used). The outcome will be analyzed as a continuous, normalized score aggregating all items. <i>Exploratory analyses will assess continuous, normalized score on the subtypes of child abuse</i>

Table 2. Caregiver level secondary impacts

Group	Indicator	Description and construction of the variable
Secondary impacts	Caregiver-adolescent relationship quality	<ul style="list-style-type: none"> The Family cohesion and communication scale is a 7-item scale scored on a 5-point Likert scale ('strongly disagree' to 'strongly agree'). These items assess communication quality in the family. The outcome will be analyzed as a continuous, normalized score aggregating all items.
	Positive Parenting and harsh discipline	<ul style="list-style-type: none"> The Positive Parenting Scale is a 9-item scale developed for conflict-affected populations, scored on a 4-point Likert scale ranging from 0 (almost never) to 3 (very frequently), range 0-27. The scale was developed to assess positive and negative parenting practices based on qualitative data with conflict-affected families (Puffer, et al, 2017). The outcome will be analyzed as a continuous, normalized score aggregating all items. The Child Abuse Screening Tool - Trial version is a 14-item scale developed by the International Society for the Prevention

		<p>of Child Abuse and Neglect (ISPCAN) (Meinck et al. 2018). The scale is meant to capture child abuse (physical, emotional, sexual, neglect) and witness violence in the past month: physical abuse (6 items), emotional abuse (6 items), and neglect (2 items). Each item scored on a binary scale: 0 (not used) or 1 (used). The outcome will be analyzed as a continuous, normalized score aggregating all items. <i>Exploratory analyses will assess continuous, normalized score on the subtypes of child abuse.</i></p>
	Mental health and stress management	<ul style="list-style-type: none"> • The Patient Health Questionnaire (PHQ-9) is a 9-item scale designed to detect and measure the severity of depression based on the Diagnostic and Statistical Manual of Mental Disorders criteria (Kroenke et al. 2001). The PHQ-9 is scored on 4-point Likert scale 0 (not at all) to 3 (nearly every day), range 0-27. The outcome will be analyzed as a continuous, normalized score aggregating all items. • The Enhanced Life Distress Inventory (ELDI) is a 14-item scale designed to capture multi-sectoral, self-perceived stress and distress across life domains (Palermo et al. 2020). The ELDI is scored on 4-point Likert scale (0–3) for concern about each item in past 7 days, 0 (none) to 3 (highly distressed), range 0-42. The outcome will be analyzed as a continuous, normalized score aggregating all items.
	Knowledge of child recruitment*	<ul style="list-style-type: none"> • <i>Child Recruitment Knowledge Index is based on 11-items developed and used as part of the GST monitoring efforts to measure direct acquisition of information conveyed in the curriculum. Each item is answered on agree or disagree response scale. The outcome is marked as exploratory, as the scale has not been validated, and will be analyzed as a continuous, normalized score aggregating all items.</i>
	Attitudes supporting child recruitment*	<ul style="list-style-type: none"> • <i>Child Recruitment Attitude Index is based on 8-items modified from tools used in the GST monitoring efforts to measure direct changes in attitudes surrounding child recruitment as taught in the curriculum. Each item is answered on a binary (agree or disagree) response scale. The outcome is marked as exploratory, as the scale has not been validated, and will be analyzed as a continuous, normalized score aggregating all items.</i>

As we pre-specify our indicators, and have a limited number of aggregated indices, with the exception of any disaggregated analysis (e.g., resilience to recruitment), we will not adjust for multiple hypothesis testing.

4. Works cited

Brock, D.M., Sarason, I.G., Sanghvi, H., et al. 2008. Perceived acceptance scale. *Journal of Social and Personal Relationships*, 14(1).

Dhar, D., Jain, T., & Jayachandran, S. 2022. Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, 112(3): 899–927.

Falb, K, Asghar K, et al, 2023. Improving family functioning and reducing violence in the home in North Kivu, Democratic Republic of Congo: a pilot cluster-randomized controlled trial of Safe at Home. *BMJ Open*, 13:e065759.

Falb, K., Fabbri, C., Bicaba, C., et al. 2026. Development of the Resistance to Recruitment (R2R) scale: A novel measure for children at risk of recruitment by armed forces and armed groups (*submitted manuscript*).

Kroenke, K., Spitzer, R.L., & Williams, J.B.W. 2001. The PHQ-9: Validity of a Brief Depression Severity Measure. *J Gen Intern Med*, 16(9): 606-613.

Lee, D.S., 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Rev. Econ. Stud*, 76 (3), 1071–1102.

Meinck, F., Boyes, M.E., Cluver, L. et al. 2018. Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. *Child Abuse & Neglect*, 82: 45-58.

Palermo, T., Hall, B.J., Cirillo, C. et al. 2020. Enhanced Life Distress Inventory: Development and Validation in two Africa countries. *Br J Health Psychol*, 25(3): 728-753.

Puffer E, Annan J, Sim AL, et al. 2017. The impact of a family skills training intervention among Burmese migrant families in Thailand: a randomized controlled trial. *PLoS One*. 12(3):e0172611.

Simes, R. J. 1986. An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 73 (3): 751–54.

Smith, B.W., Dalen, J., Wiggins, K., et al. 2008. The brief resilience scale: Assessing the ability to bounce back. *Inter J Behav Med*, 15(3): 194-200.

Snyder, C.R. Hoza, B., Pelham, W.E., et al. 1997. The development and validation of the Children's Hope Scale. *J Pediatric Psychology* 22(3): 399-421.