Evaluation of a school mental health package in Uganda

Analysis Plan, Version 1.0

June 19, 2026

SECTION 1: ADMINISTRATIVE INFORMATION

1. Title and registration

Evaluation of a school mental health package in Uganda. This trial was registered on ISRCTN on June 1, 2025: ISRCTN34179105.

2. SAP Version

1.0, June 25, 2026

3. Protocol version

Version 3.0, Dated 26 June 2025

4. SAP revisions

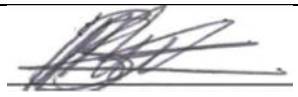
None

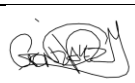
5. Roles and responsibility

Statistician: Dr Sebastian Kurten
Investigators: Prof Mark JD Jordans (PI), Sandra Agondeze, Dr Byamah Mutamba, Dr Myrthe van den Broek, Anthony Guevara, Dr. Gabriela Koppenol-Gonzalez, Clotilde Mahe, Prof. Wendy Janssens, Prof. Brandon Kohrt

6. Signatures of investigators

Reviewed by all listed authors and signed by the following.

Name	Signature	Date
Prof Mark JD Jordans		24-06-2026
Sandra Agondeze		24-06-2026
Byamah Mutamba		23-06-2026
Myrthe van den Broek		22-06-2026
Anthony Guevara		22-06-2026
Sebastian Kurten		24-6-2026

Gabriela Koppenol-Gonzalez		23-06-2026
Clotilde Mahe		22-06-2026
Prof Wendy Janssens		21-06-2026
Prof Brandon Kohrt		25-06-2026

SECTION 2: INTRODUCTION

This statistical analysis plan describes procedures for analyzing data from a randomized waitlist-controlled cluster trial of a Mental Health and Psychosocial Support (MHPSS) package in Uganda. Relevant study design details are provided below followed by a detailed description of the statistical analysis approach. This analysis plan only covers the quantitative data analyses and excludes the qualitative analysis approach. The statistician and PI will remain masked to study conditions until primary analyses have been completed.

The analyses described below will inform the development of one scientific manuscript:

- Effectiveness of a multi-component mental health care system for adolescents in refugee settings

7. Background and rationale

Children and adolescents in humanitarian and conflict-affected settings face substantial and interrelated threats to their mental health and psychosocial wellbeing. Exposure to violence, displacement, loss, and chronic instability is associated with increased emotional and behavioral difficulties, with potential consequences for development, educational attainment, and longer-term life outcomes. These impacts are shaped not only at the individual level but across adolescents' social ecology. Caregivers experiencing high levels of stress may have reduced capacity to provide consistent and responsive care; teachers often work under conditions that undermine their wellbeing and ability to support students; and communities may face disrupted social cohesion, increased violence, and reliance on harmful coping strategies. Stigma further compounds these challenges by limiting help-seeking and access to care. Together, this evidence underscores that adolescents' mental health in such settings is embedded within, and contingent on, broader family, school, and community systems.

Despite growing availability of evidence-based mental health and psychosocial support (MHPSS) interventions, current responses remain fragmented, under-resourced, and insufficiently integrated across sectors. Services are frequently delivered in isolation, with limited coordination between health, education, and community platforms, constraining coverage, continuity of care, and effectiveness. Interventions often prioritize the individual child, with insufficient attention to the wider social determinants of mental health, including the wellbeing of caregivers, teachers, and community structures. As a result, both reach and effectiveness are limited, particularly in low-resource and conflict-affected settings. There is a paucity of rigorous evidence on comprehensive, multi-level approaches that operate across ecological domains and systems of care. This study aims to address this gap by evaluating an integrated, ecologically grounded MHPSS care system that targets adolescents, caregivers, teachers, and communities, with the objective of improving psychosocial wellbeing and strengthening supportive environments at scale.

8. Objectives

The overall aim of the study is to evaluate the impact of a mental health care system on improving
mental health of adolescents in refugee settlements, with the following guiding
research question and hypothesis:

- What is the effectiveness of a multi-level multi-component care system on the wellbeing amongst adolescents (aged 11-16), and/or depression, anxiety, and internalizing symptoms amongst adolescents (aged 11-16 years) with elevated levels of symptoms?

Hypothesis: Adolescents in the experimental arm (i.e. mental health care system), regardless of whether and what services they received, will have (i) a higher level of wellbeing and/or
(ii) lower levels of depression, anxiety, or internalizing symptoms amongst those with elevated levels of symptoms at baseline (above validated cut-off), over a period of 12 months compared with adolescents in the control arm (waitlist control).

SECTION 3: STUDY METHODS

9. Trial design

The study is conducted across three refugee settlements in western Uganda — Kyangwali (~150,000 refugees), Kyaka II (~135,000), and Nakivale (~266,000) — hosting displaced persons primarily from the Democratic Republic of Congo (DRC), where adults work predominantly as crop farmers or farm laborers (Government of Uganda OPM & UNHCR, 2025). Uganda was selected due to its large refugee population, our extensive preliminary work in the region, and its relevance as a low-resource setting: it is one of the world's largest refugee-hosting nations, with approximately 1.4 million refugees, over 60% of whom are children.

The study uses a pragmatic parallel two-arm cluster Randomized Controlled Trial (cRCT), randomizing 18 schools (clusters), across the three settlements, 1:1 (stratified by school size) to either a multi-component mental health care system (experimental) or waitlist control. Within each zone, one school was selected, and a random sample of adolescents aged 11–16 was drawn from each selected school for data collection. Sampled adolescents and their caregivers were interviewed at baseline before the roll-out of the intervention in the experimental arm and followed up at 6 and 12 months after completion of the intervention in the experimental arm but before roll-out in the waitlist control arm.

The program and its separate interventions are described in detail in the study protocol (ISRCTN34179105). The design aims to mimic a naturalistic service delivery framework: while the program was offered at the zone and school level, not all households in treatment zones and not all classes in treatment schools received the full set of services. Thus, sampled adolescents and caregivers in the experimental arm may have received some, all, or no services. Moreover, non-sampled adolescents and caregivers in experimental schools/zones may also have accessed services.

As a result, sampled adolescents in the experimental arm may have benefited both directly (receiving services themselves) and/or indirectly (through peers and caregivers).

We follow intention-to-treat as the primary analysis framework. A cRCT was chosen to prevent contamination across adolescents, caregivers, and school staff.

10. Randomization

The randomization was done in two stages. The first randomization stage consisted of 5 steps:
Step 1: From a pool of 24 identified eligible schools, 19 schools were selected using a reproducible random selection procedure in R. Thereby, we randomly selected one school when there were two schools within one district. Then, we removed the smallest school to reach 18 schools.
Step 2: The 18 schools were stratified by number of learners, and we selected 9 pairs of schools that were most similar in terms of size.
Step 3: The two schools per pair were randomly allocated to one of the two study arms (Arm A and Arm B) using a second reproducible R script. Both scripts are archived and reproducible.
Step 4: To determine which arm served as the intervention and which as the control, a coin-flip procedure was employed to maintain allocation of concealment. Prior to the coin flip, the study coordinator privately documented her decision linking coin outcomes to arm assignments (e.g., heads = Arm A intervention vs. heads = Arm B intervention) and sealed this in an envelope, which was stored confidentially by the PI and will be opened only after data analysis is complete.
Step 5: A co-investigator then flipped the coin and communicated the result to the scientific coordinator, with the process visually documented. This two-step procedure ensured that the scientific coordinator could identify the intervention and control arms for operational purposes, while all investigators will remain masked to the allocation until the analysis is finalized.
For stage 2 randomization, we randomly selected 126 female and 126 male students from a full list of students in each of the 18 schools. In case students were unavailable or unwilling to participate in the data collection, additional students were randomly sampled to serve as replacement. This was done using a reproducible R-script.

11. Sample size

Sample size calculations were based on two primary outcomes: adolescent wellbeing measured by the Stirling Scale and psychological distress (i.e. depression, anxiety, and internalizing symptoms) measured by the MMAPP. The trial is considered successful if a significant intervention effect is observed on (i) improved wellbeing in the total study sample and/or (ii) a reduction in psychological distress among adolescents with elevated symptoms at baseline (score >10 on the MMAPP). For the Stirling Wellbeing Scale, we assume we can detect an effect size of 0.33. For the MMAPP, we focus on adolescents with clinically relevant distress (scores >10), and can assume to detect an effect size of 0.34. Based on pilot-data, we assumed that at least 104 adolescents (around 60%) per cluster would score above that threshold. To account for the multiple primary outcomes, the Benjamini-Yekutieli correction was applied. An intraclass correlation coefficient of 0.04 was assumed based on published

literature, and a desired statistical power of 80% was specified. The calculations were conducted using the CRTSize package in R.
The sample size calculations resulted in 18 clusters (schools) allocated at a 1:1 ratio, with 9 schools per arm. Accounting for an assumed 30% loss to follow-up, a minimum of 252 adolescents per cluster was recruited, corresponding to an effective analytical sample of 176 adolescents per cluster. This resulted in a target sample size of 4,536 adolescents across all 18 schools.
The original sample size calculation applied the Benjamini-Yekutieli correction ($\alpha = .0333$ for the primary outcome with the lower p-value). The SAP specifies the Hochberg procedure for the final analysis (thresholds .025/.05).

12. Framework

The objective of the study has a superiority hypothesis testing framework. It aims to investigate the superiority of the mental health care arm relative to the waitlist control arm on the primary outcomes.

13. Statistical interim analyses and stopping guidance

No interim analyses of study outcomes were planned or conducted, except for data quality control purposes. The Data Safety Management Committee reviewed adverse events throughout the study implementation and were responsible for determining whether any actions were required to reduce risk to study participants.

14. Timing of final analysis

All analyses will be conducted after completion of data collection for the trial. The investigators (except study co-ordinator) statistician and PI will be masked to study condition during data collection, cleaning, and the primary effectiveness analyses. Secondary analyses examining mediational pathways, spill-over effects on other domains of well-being and cost-effectiveness analysis will be covered in a separate SAP and will be conducted separately from the effectiveness analyses presented in this SAP.

15. Timing of outcome assessments

Data used for this study are collected by research assistants at baseline (May 2025), midline (November 2025), and endline (June 2026), from adolescents and their caregivers. Caregivers were only interviewed at baseline and endline. The midline assessment consisted of a reduced set of measures and was only administered to adolescents.

SECTION 4: STATISTICAL PRINCIPLES

16. Level of statistical significance

A significance level of alpha = 0.05 will be applied, but it will be adjusted for testing two primary outcomes.

17. Adjustments for multiplicity

Given that the trial has two primary outcomes, we will apply the Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) to control the false discovery rate. This procedure is valid under the assumption of positive regression dependency between tests (Benjamini & Yekutieli, 2001), which is considered plausible given that the test-statistics of both outcomes are expected to be positively correlated.

It was also considered to use the Hochberg-correction as a mean of correcting for the family-wise error rate (Sarkar 2008). However, for the case of two tests, both correction methods are statistically equivalent, meaning that they produce identical cut-off values. Therefore, the multiplicity adjustment controls for both, false discovery rate, and family-wise error rate.

Under both procedures, both null hypotheses are rejected if both p-values are ≤0.05; otherwise, the hypothesis with the smaller p-value is rejected if that p-value is ≤0.025.

18. Confidence intervals

Results will be presented as coefficients and 95% CIs.

19. Adherence and protocol deviations

Not all study participants in the treatment arm will be receiving intervention(s). For those that do, participant adherence to the intervention will be measured using intervention session attendance records (for each intervention separately). Completion of the intervention will be defined as having attended >75% of the sessions. In case of multiple interventions, they will need to at least attend 75% of each intervention. We will report the mean (SD) and median (IQR) number of sessions attended in the intervention group.

Facilitator fidelity to the intervention will be measured using a fidelity checklist (separately for each of the interventions, all measured through in-session observations of facilitators providing sessions). We will report the mean (SD) and median (IQR) scores on the fidelity checklist as an indicator of trial implementation quality. Facilitators competency will be measured using the EQUIP competency assessment tools (separately for each of the interventions, all measured through in-session observations of facilitators providing sessions, as well as in-training observation). We will report the proportion of facilitators demonstrating adequate competency (levels 3 and 4 on the EQUIP tools) across observations, as an indicator of trial implementation quality.

Any major deviations from the protocol will be reported in the main outcomes paper.

20. Analysis populations

The primary analysis will be conducted on the intent-to-treat sample, meaning that all participants will be included and analysed as randomized. No enrolled participants and/or clusters were excluded from the main analytic sample.

SECTION 5: TRIAL POPULATION

21. Screening data

There were no diagnostic no clinical screening criteria for admission into the study. Participation was open to all adolescents enrolled at participating schools, regardless of mental health status or symptom levels.

22. Eligibility

The study aimed to enrol 4536 adolescents and their caregivers across 18 schools in refugee settlements in Western Uganda. Schools were randomized to the intervention or waitlist control arm with 1:1 allocation. Randomization at school-level ensured that all adolescents from a given school were allocated to the same study arm. Schools were eligible for the trial if they:

(i) were willing and permissioned to participate;
(ii) had sufficient learners between 11-16 (n > 300);
(iii) had pupil lists or registers available;
(iv) were > 5 km from a school in an adjoining clusters/zone that had been enrolled in the study (in that case, we used simple random selection to determine which school gets excluded)

In any selected school, we performed a random selection of 252 participants. Adolescents were deemed eligible if: (i) they were between 11 and 16 years; (ii) spoke one of the study languages (Kinyabwisha, Kiswahili, Runyankole, and Runyoro-Rutooro); (iii) attended the selected schools at the time of enrolment; and (iv) the adolescents and their caregivers had to provide active consent.

23. Recruitment

Participants were recruited through schools. Research assistants asked the randomly selected students at a given school to participate in the study. Students were approached until the desired target size of 252 students per cluster was achieved.

24. Withdrawal and loss to follow-up

Data on withdrawal and loss to follow-up will be reported in the CONSORT diagram in the main outcomes paper, including reported reasons for withdrawal, if known (see Figure 1).

The frequency of intercurrent events (school transfer, school dropout, relocation, and cross-arm access to MHPSS services, where measurable) will be reported descriptively by arm.

Stage 1 — Enrolment and randomization of clusters (schools)

↓

Schools assessed for eligibility: n = 24

↓

Excluded: n = 6 schools
Not selected for inclusion (random selection of 19 of 24 eligible schools among schools in the same district, and exclusion of smallest school)

↓

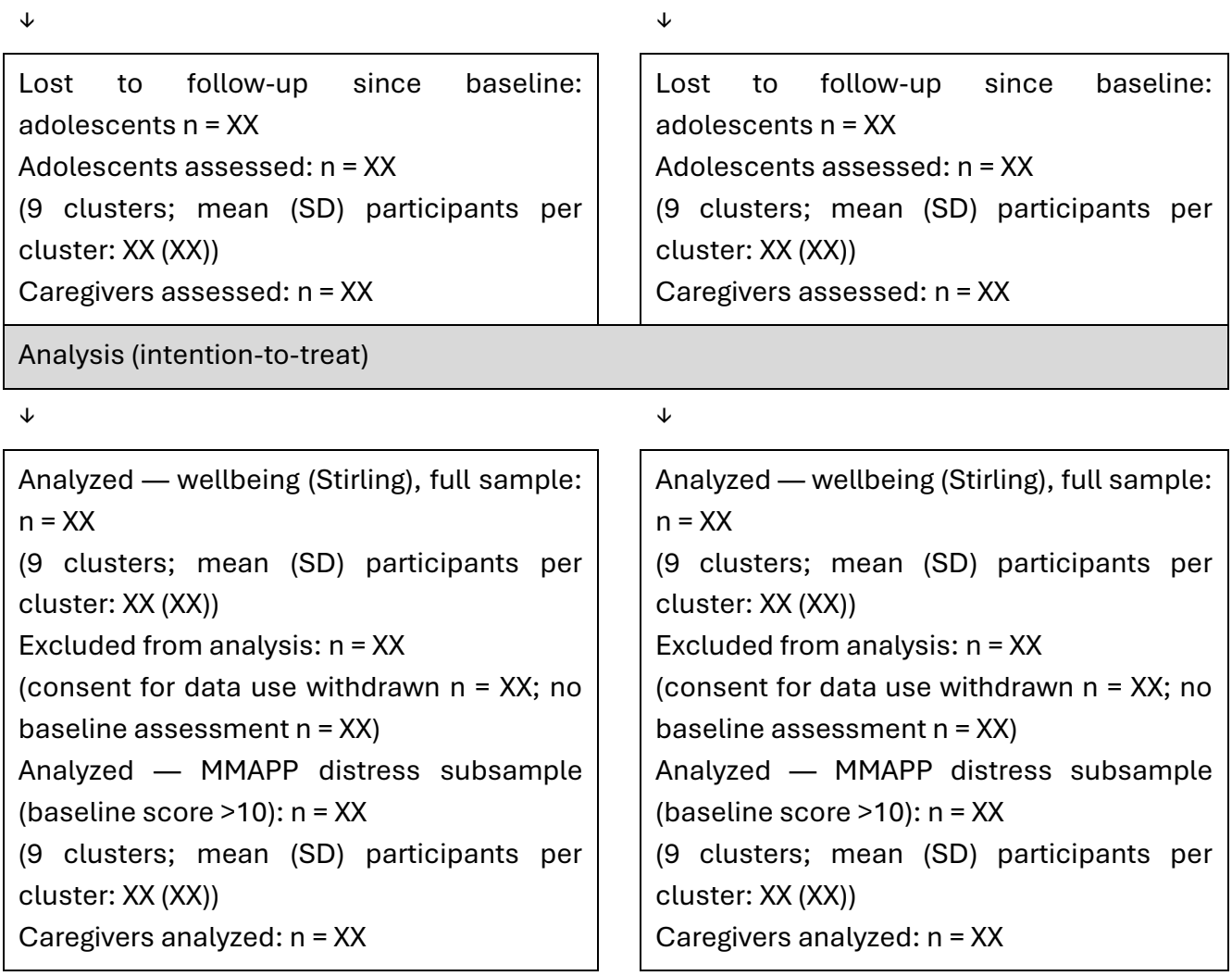
Stratification: 18 schools matched into 9 pairs by number of students

↓

Randomized: n = 18 clusters (9 pairs; 1:1 allocation; concealed coin-flip procedure for arm assignment)

↓ ↓

MHPSS intervention arm Allocated: 9 clusters	Waitlist control arm Allocated: 9 clusters
Stage 2 — Sampling of participants within clusters	
Adolescents enrolled: n = XX (target 252 per cluster; 9 clusters; mean (SD) participants per cluster: XX (XX)) Caregivers enrolled (linked to child): n = XX	Adolescents enrolled: n = XX (target 252 per cluster; 9 clusters; mean (SD) participants per cluster: XX (XX)) Caregivers enrolled (linked to child): n = XX
Baseline assessment (May 2025)	
Adolescents assessed: n = XX (9 clusters; mean (SD) participants per cluster: XX (XX)) Caregivers assessed: n = XX	Adolescents assessed: n = XX (9 clusters; mean (SD) participants per cluster: XX (XX)) Caregivers assessed: n = XX
Midline assessment (Nov 2025) — adolescents only, abbreviated survey	
Lost to follow-up: n = XX Adolescents assessed: n = XX (9 clusters; mean (SD) participants per cluster: XX (XX))	Lost to follow-up: n = XX Adolescents assessed: n = XX (9 clusters; mean (SD) participants per cluster: XX (XX))
Endline assessment (Jun 2026)	



Note. Diagram follows the CONSORT 2010 extension for cluster randomized trials (Campbell et al., 2012). All analyses follow an intention-to-treat framework; participants with a baseline assessment but missing endline data are retained in the analysis via multiple imputation and are reported under 'lost to follow-up', not under 'excluded from analysis'. 'Excluded from analysis' is reserved for participants who withdrew consent for data use or had no baseline assessment. The MMAPP distress subsample comprises adolescents scoring >10 on the MMAPP at observed baseline and forms the analytic sample for the psychological distress dual primary outcome only. Caregivers were assessed at baseline and endline only.

25a. List of baseline socio-demographic characteristics to be summarized (see appendix)

- Location (school/cluster membership)
- Child age
- Child sex
- Child nationality
- Child language
- Caregiver age
- Caregiver gender
- Caregiver nationality
- Caregiver language

- Relationship of caregiver to the index child

25b. How baseline characteristics will be descriptively summarized

We will describe the distribution of demographic characteristics as well as primary and secondary outcomes at baseline using measures of central tendency in the overall sample and stratified by intervention group (see appendix). Baseline comparisons will be descriptive, and no statistical significance tests or confidence intervals will be calculated for the difference between randomized groups on any participant level baseline variables. The randomization of participants to intervention groups means that any imbalance over all measured and unmeasured baseline characteristics is by definition due to chance.

SECTION 6: ANALYSIS

26a. Estimands

The primary estimand targets the effect of allocation to the MHPSS care system, regardless of services received. For the wellbeing outcome, the population is all randomized adolescents; for the psychological distress outcome, the population is adolescents scoring >10 on the MMAPP at baseline. The variable is the respective total score at endline, and the population-level summary measure is the covariate-adjusted mean difference between arms. Intercurrent events (e.g., school transfer, school dropout, relocation outside the settlement, and access to MHPSS services outside the allocated arm) will be handled under a treatment policy strategy: all adolescents will be analyzed according to the randomized allocation of their school at baseline and followed up wherever feasible. Where intercurrent events result in missing endline assessments, these will be handled under the missing data strategy in Section 28.

26b. Outcome definitions

The primary outcomes for this study are psychological distress (i.e. depression, anxiety, and internalizing symptoms) and psychological wellbeing. Secondary outcomes on child-level are behavioral problems, mental health related risk factors, health-related quality of life, friendship, substance abuse and stigma. On caregiver level, secondary outcomes include parenting practices, caregiver wellbeing, and caregiver psychological distress.

The measures – including outcomes, and implementation variables - and how they are operationalized and measured are described below.

Table 1. Study outcome measures and timepoints

Measures	Baseline	Midline	Endline
Primary Outcomes (adolescent)			
Psychological distress (MMAPP)	X	X	X
Psychological wellbeing (Stirling scale)	X	X	X
Secondary Outcomes (adolescent)			
Behavioral problems (DBIS)	X	X	X
Mental health risk factors (IDEA-RS)	X	X	X
Health-related quality of life (KIDSCREEN-10)	X		X
Stigma (EDS)	X		X
Hope (Children's Hope Scale)	X	X	X
Social connectedness (SCS)	X		X
Friendship	X		X
Alcohol & substance use	X		X

Measures	Baseline	Midline	Endline
Secondary Outcomes (caregiver)			
Parenting (BPQ)	X		X
Psychological distress (K6)	X		X
Psychosocial wellbeing (WEMWBS)	X		X
MHPSS Participation TeamUp participation (child)			X
Ease participation (child)			X
Ease participation (caregiver)			X
BeThere participation (caregiver)			X

Primary outcomes

- **Psychological distress:** Total score on a 25-item scale (see Appendix), the UNICEF Measuring Mental Health of Adolescents and Young People at a Population Level (MMAPP). It enables the assessment of anxiety and depression symptoms, using a 4-point response scale (0–3; never, sometimes, often, always). The outcome was assessed at baseline, midline, and endline.
- **Wellbeing:** Total score on the Stirling wellbeing scale. It consists of 15 items, using five response options (0-4; Never, not much of the time, some of the time, quite a lot of the time, all of the time). The outcome was assessed at baseline, midline, and endline. The total score combines items on emotional state (items 9, 10, 11, 12, 14, 15) and positive outlook (1, 3, 4, 5, 6, 8). The remaining three items are indicators for social desirability and should not be included.

Secondary outcomes (adolescent)

- **Behavioral problems (child-level):** Total score on the 8-item version of the Disruptive Behavior International Scale - Nepal version (DBIS) (Burkey, 2018). Response options follow a Likert scale; Never (0), Sometimes (1), Often (2), Very Often (3). The outcome was assessed at baseline, midline, and endline.
- **Mental health related risk factors (child-level):** The Identifying Depression Early in Adolescents (IDEA-RS) tool will be used to determine and measure burden of risk factors for future development of depression. IDEA-RS includes 11 domains (gender, minority group, school failure, running away from home, getting into fights, relationship with peers, relationship with mother, relationship with father, relationship between parents, maltreatment, and drug use). The IDEA-RS total is derived from items on other scales as well as six unique items, with three yes/no questions and three questions using a 3-point scale (Never, Sometimes, Often). The outcome was assessed at baseline, midline, and endline.
- **Health-related quality of life (child-level):** Total score on the KIDSCREEN-10. It is a 10-item scale, assessing health-related Quality of Life. There are five response options: (0) Never, (1) Seldom, (2) Quite often, (3) Often, and (4) Always. The outcome was assessed at baseline and endline.

- **Stigma (child-level):** The 8-item Everyday Discrimination Scale (EDS) will be used to asses stigma. It is a six-point scale (0-5, Never, Less than once a year, A few times a year, A few times a month, At least once a week, Almost every day). The outcome was assessed at baseline and endline.
- **Hope scale (child-level):** Total score on the 6-item Children's Hope Scale (Snyder, 1997), measuring hopeful thinking and goal-directed beliefs. It is a six-point scale (0–5; None of the time, A little of the time, Some of the time, A lot of the time, Most of the time, All the time). Higher scores indicate greater hope and goal-directed behavior. The outcome was assessed at baseline, midline, and endline.
- **Social connectedness (child-level):** Total score on the 8-item Social Connectedness Scale, assessing the degree to which youth feel connected to others in their social environment. It is a five-point scale (0–4; Strongly disagree, Disagree, Slightly disagree, Agree, Strongly agree). The outcome was assessed at baseline and endline.
- **Friendship (child-level)**: Friendship is assessed with three items covering number of close friends, proportion who are refugees, and frequency of socializing outside of school. The outcome was assessed at baseline and endline. We will investigate those items separately in an exploratory analysis.
- **Alcohol use (child-level):** Three items assessing alcohol use, covering own alcohol consumption in the past three months, whether close friends consume alcohol, and whether close friends have been drunk in the past three months. The outcome was assessed at baseline and endline. We will investigate those items separately in an exploratory analysis.

Secondary outcomes (caregiver)

- **Parenting (caregiver-level):** Total score on a 23-item parenting scale developed to assess change in parenting. The total score combines subscales assessing parental warmth and sensitivity (17 items) and harsh parenting (6 items). The outcome was assessed at baseline and endline.
- **Caregiver psychological distress (caregiver-level):** Total score on the 6-item Kessler psychological distress scale. It is a five point Likert scale (0-4, None of the time, A little of the time, Some of the time, Most of the time, All the time) Caregiver stress was assessed at baseline and endline.
- **Caregiver psychosocial wellbeing (caregiver-level):** Total score on the 14-item Warwick-Edinburgh Mental Wellbeing Scale. It is a five point Likert scale (0-4, Never, Rarely, Sometimes, Often, Always) Caregiver psychosocial wellbeing was assessed at baseline and endline.

Mental health care participation indicators

The mental health care intervention system consists of several interventions. We will report the proportion of the research sample (treatment arm) that fall in any of the below categories:

1. TeamUp alone
2. EASE alone
3. BeThere alone
4. TeamUp and EASE
5. TeamUp and BeThere

6. TeamUp, EASE and BeThere
7. No services

27a. Analysis methods

Data Management

We will prepare both a long and wide final dataset. The long dataset will include a separate observation for each assessment period. The wide dataset will include one observation per individual with variables for each assessment period. The final dataset will include the following variables:

- School identifier
- Household identifier
- Individual identifier (caregivers, child)
- Allocation [masked: will be coded 0 or 1 without a label until primary analyses are complete]
- Assessment number/time point
- Caregiver age
- Child age
- Caregiver gender
- Child gender
- Child psychological distress items and total score (MMAPP)
- Child psychological wellbeing items and total score (Stirling scale)
- Child behavioral problems items and total score (DBIS)
- Child mental health risk factors items and total score (IDEA-RS)
- Child health-related quality of life items and total score (KIDSCREEN-10)
- Child stigma items and total score (EDS)
- Child hope items and total score (Children's Hope Scale)
- Child social connectedness items and total score (SCS)
- Child friendship items
- Child alcohol and substance use items
- Caregiver-reported parenting items and total score (BPQ)
- Caregiver psychological distress items and total score (K6)
- Caregiver psychosocial wellbeing items and total score (WEMWBS)
- TeamUp participation Child
- EASE participation child
- EASE participation caregiver
- BeThere participation caregiver

We will review the data and perform the following quality checks prior to beginning the analysis: 1) identify and evaluate any missing data; 2) review the distribution of all variables to identify any outliers or potential data entry errors, including reviewing the distribution of outcomes at each time point; and 3) search for any data irregularities. Any identified issues will be reviewed and reconciled with the research team. Once the dataset has been cleaned, we will proceed with creating and coding the

variables required to answer the primary and secondary objectives. The variables used in the final analysis are described in table 1.

Descriptive statistics

The distribution of baseline characteristics will be reported for the overall sample and stratified by study arm. For each primary and secondary outcome, a summary of results will be presented by trial arm for each time point. Appropriate summary statistics will be applied to describe demographic and baseline characteristics (appendix 2): mean and standard deviation for all symmetric (non-skewed) distributed measures; median, 25th and 75th quartiles for skewed distributions. QQ plots and histograms will be used to assess data distributions of continuous measures. Categorical outcomes will be described using both numbers and proportions (percentage). The following variables will be reported for the overall sample and by arm: location, child age, child sex, child nationality, child language, caregiver age, caregiver sex, caregiver nationality, caregiver language, and caregiver relationship to child.

Loss to follow-up, active withdrawals from the trial (e.g., revoking consent), and the prevalence of serious adverse events will be reported at endline.

Primary Effectiveness Analysis

The analyses outlined in this strategy will be pragmatic, based on ITT, and will utilize all available follow-up data from all randomized participants. The trial statistician will be masked to treatment allocation. The main statistical analyses will establish the effectiveness of receiving MHPSS over waitlist control by comparing the mean difference of total psychological distress (MMAPP) and wellbeing (Stirling) scores at endline (see appendix 3).

Model assumptions will be checked using residual diagnostics (QQ plots, residuals vs. fitted values). In case of non-convergence, the optimizer will be varied first; if non-convergence persists, the cluster-level covariate will be removed.

Primary outcome

The dual primary outcomes are the total scores of the MMAPP and the Stirling Wellbeing Scale at endline.

We will fit two two-level linear mixed models (Gaussian distribution, identity link) to analyze the dual primary outcomes, with observations at the individual level (level 1) and a random intercept for school (level 2) to account for the clustering of adolescents within schools (Billot et al. 2024). Both models adjust for the respective baseline outcome score, child sex, and age. To account for the stratified randomization design, in which schools were matched by size prior to allocation, school size (the stratification variable, derived from total school enrolment) is included as a cluster-level covariate.

Covariate adjustment serves to improve precision and to adjust for potential chance imbalance arising from the small number of clusters.

The first model includes the total MMAPP score as the dependent variable and is fitted on the subsample of adolescents scoring above the clinically validated threshold of >10 at baseline. The cut-off was clinically validated as part of this project; validation results will be published separately. The second model includes the total Stirling Wellbeing score as the dependent variable and is fitted on the full analytical sample. In both models, treatment allocation is modelled as a binary fixed effect, and its coefficient is the primary estimate of interest, reported as a mean difference with 95% confidence interval, p value, and effect size (Cohen's d).

The intraclass correlation coefficient for each primary outcome will be estimated from the fitted model and reported.

Secondary outcomes

Secondary outcomes as listed will be assessed using a similar methodology to the primary outcomes, using generalized linear mixed models with the appropriate link function and distribution for each outcome. Normal outcomes will use an identity link function, binary variables will use a logit link function, and Poisson (count) variables will use a log link function. For all secondary outcomes, the analysis will compare randomized arms at endline adjusted for baseline scores, child sex, and age consistent with the approach used for the primary outcomes. School size will also be included as cluster-level covariate in the models.

Caregiver secondary outcomes (psychological distress [K6], psychosocial wellbeing [WEMWBS], and parenting practices [BPQ warm and harsh subscales]) will be analyzed using the same generalized linear mixed modelling approach as the adolescent secondary outcomes. Each model will include endline score as the outcome, treatment allocation and baseline score as explanatory variables, and a random intercept for school (cluster) to account for the clustering of caregivers within schools. As caregivers were assessed at baseline and endline only, no midline data will be available for longitudinal sensitivity analyses of caregiver outcomes.

Planned sensitivity analysis

In addition to the primary intention-to-treat analysis, we will also conduct a per protocol analysis. The per protocol analysis will remove data of those who did not complete both baseline and endline assessments.

Subgroup analyses will be conducted for the two primary outcomes, stratified by child sex, age group (11–13 vs. 14–16 years), and nationality (Ugandan vs. non-Ugandan). Subgroup effects will be examined by introducing a treatment-by-subgroup interaction term into the primary outcome model. These analyses are exploratory and hypothesis-generating in nature; the trial was not powered to detect subgroup differences and findings should be interpreted with caution. No multiplicity correction

will be applied to subgroup interaction tests, and results will be reported as interaction coefficients with 95% confidence intervals.

Further, we will conduct a longitudinal analysis for both primary and secondary outcomes where midline data are available, fitting a three-timepoint repeated measures mixed model (baseline, midline, endline) to assess whether intervention effect trajectories are consistent with the primary endline findings.

Adolescents were sampled at the individual level from school registers, and household membership was not a sampling unit; however, siblings attending the same school may have been sampled into the study. We will report the number and proportion of participants sharing a household identifier. Because any such clustering is nested within schools and expected to be rare, the primary model does not include a household-level random effect. As a sensitivity analysis, if more than 5% of participants share a household with another participant, the primary models will be re-fitted with an additional random intercept for household, and results will be compared with the primary analysis.

As an exploratory analysis, we will refit the primary MMAPP model on the full analytical sample to estimate the intervention effect across all randomized adolescents, rather than only those above the baseline distress threshold. The model specification will be identical to the primary analysis (random intercept for school; adjustment for baseline MMAPP score, child sex, age, and school size), with the treatment coefficient reported as a mean difference with 95% confidence interval, p value, and Cohen's d.

28. Missing data

For questionnaire outcome measures where there are published methods for dealing with missing items, these will be applied. Otherwise, we will prorate missing items only when there are no more than 20% missing items (i.e. for a ten-item questionnaire, prorate only where one or two items are missing) by replacing the missing item values with the median value of the complete items for each individual.

In accordance with guidelines on handling missing data in ITT analysis (Twisk et al. 2020), multiple imputation will be used to handle missing outcome data for participants who completed baseline assessment but not endline assessment. Missing data will be assumed to be missing at random (MAR). Multiple imputation by chained equations (MICE) will be used, with 50 imputed datasets.

29. Additional analyses

Not applicable

30. Harms

Serious Adverse Events—harm to self, harm to others, family violence, and other events identified by the study team in the course of the study have been carefully monitored and reported on and reviewed by the Data Safety Management Board. Number and types of events will be reported. Referral and follow-up of individual cases will be done as indicated.

31. Statistical software

Data will be imported into R for data management and analysis.

32. References

Billot, L., Copas, A., Leyrat, C., Forbes, A., & Turner, E. L. (2024). How should a cluster randomized trial be analyzed? Journal of Epidemiology and Population Health, 72(1), 202196. https://doi.org/10.1016/j.jeph.2024.202196

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29(4). https://doi.org/10.1214/aos/1013699998

Burkey, M. D., Adhikari, R. P., Ghimire, L., Kohrt, B. A., Wissow, L. S., Luitel, N. P., Haroz, E. E., & Jordans, M. J. D. (2018). Validation of a cross-cultural instrument for child behavior problems: The Disruptive Behavior International Scale – Nepal version. BMC Psychology, 6(1), 51. https://doi.org/10.1186/s40359-018-0262-z

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., Williamson, P., Altman, D. G., Montgomery, A., Smyth, E. T., & Loder, E. (2017). Guidelines for the content of statistical analysis plans in clinical trials. JAMA, 318(23), 2337–2343. https://doi.org/10.1001/jama.2017.18556

Government of Uganda, Office of the Prime Minister, & United Nations High Commissioner for Refugees. (2025, May 31). Uganda population statistics dashboard – 31 May 2025. UNHCR Operational Data Portal. https://data.unhcr.org/en/documents/details/116727

Sarkar, S. K. (2008). Generalizing Simes' test and Hochberg's stepup procedure. The Annals of Statistics, 36(1). https://doi.org/10.1214/009053607000000550

Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., Highberger, L., Rubinstein, H., & Stahl, K. J. (1997). The development and validation of the Children's Hope Scale. Journal of Pediatric Psychology, 22(3), 399–421. https://doi.org/10.1093/jpepsy/22.3.399

Twisk, J. W. R., Rijnhart, J. J. M., Hoekstra, T., Schuster, N. A., ter Wee, M. M., & Heymans, M. W. (2020). Intention-to-treat analysis when only a baseline value is available. *Contemporary Clinical Trials Communications*, *20*, 100684. https://doi.org/10.1016/j.conctc.2020.100684

Wright, N., Ivers, N., Eldridge, S., Taljaard, M., & Bremner, S. (2015). A review of the use of covariates in cluster randomized trials uncovers marked discrepancies between guidance and practice. Journal of Clinical Epidemiology, 68(6), 603–609. https://doi.org/10.1016/j.jclinepi.2014.12.006

Appendix: Codebook & Table shells

1. Study variables (Section 25a) - Abbreviated Codebook

This abbreviated codebook lists variables from the baseline child questionnaire that are referenced in the Statistical Analysis Plan. Variables are grouped by their role in the analysis (administrative, demographics, primary outcomes, secondary outcomes, mediators, exploratory outcomes). Item-level variables are shown where individual items are used in quality checks or subgroup analyses; computed sum scores are flagged with square brackets []. All missing value codes are consistent: 88 = Prefer not to answer (PNTA); 99 = Do not know (DK).

I. Administrative and identification variables

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
A. Administrative / identification variables				
school	School identifier (cluster)	Nominal	1–18 (18 schools)	Cluster-level random effect; primary analysis unit
code_child	Child unique identifier	Nominal	Free text	Links repeated assessments; ITT tracking
date	Interview date	Date	DD/MM/YYYY	Assessment timing / protocol deviation checks
id_ra_code	Research assistant code	Nominal	Free text	RA masking checks (Section S)
language_interview	Interview language	Nominal	1 Kinyabwisha; 2 Kiswahili; 3 Runyankole; 4 Runyoro-Rutooro; 5 None (ineligible)	Eligibility; language subgroup analyses
cg_consentforchild	Caregiver consent for child participation	Binary	0 No; 1 Yes	Eligibility filter

II. Demographic variables (baseline characteristics)

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
B. Demographic variables (baseline characteristics)				
child_age	Child age in years	Continuous	Integer (target: 11–16)	Baseline characteristic; subgroup analysis (11–13 vs 14–16)
child_gender	Child gender (observed by RA)	Nominal	1 Male; 2 Female	Baseline characteristic; gender subgroup analysis
child_nationality	Child nationality	Nominal	1 Ugandan; 2 Congolese; 3 South Sudanese; 4 Burundian; 5 Rwandese; 6 Other; 99 Unknown	Baseline characteristic; nationality subgroup analysis (Ugandan vs non-Ugandan)
Languages	Language(s) spoken at home	Multi-select nominal	1 Kinyabwisha; 2 Kiswahili; 3 Runyankole; 4 Runyoro/Rutooro; 5 Other	Baseline characteristic (descriptive)

III. Primary outcomes

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
D. Primary outcome — Psychological Wellbeing (Stirling Children's Wellbeing Scale)				
stirling_1 – stirling_15 (Items 2, 7, and 13 do not count towards total score)	12 items assessing positive wellbeing over past two weeks (e.g., 'Did you think good things will happen in your life?')	Ordinal (per item)	0 Never; 1 Not much of the time; 2 Some of the time; 3 Quite a lot; 4 All of the time; 88 PNTA; 99 DK	Primary outcome (Manuscript 1 & 2). Sum score = total Stirling. Baseline covariate in primary model.
[stirling_total]	Total Stirling score (computed)	Continuous	Sum of 12 stirling items (range 0–48; higher = better wellbeing)	Dual primary outcome; endline score compared between arms adjusted for baseline
E. Primary outcome — Psychological Distress (MMAPP)				
mmapp_1 – mmapp_25	25 symptom items across depression, anxiety, somatic domains over past two weeks (e.g., 'how often have you been feeling very sad or depressed?')	Ordinal (per item)	0 Never; 1 Sometimes; 2 Often; 3 Always; 88 PNTA; 99 DK	Primary outcome (Manuscript 1 & 2). Sum score = total MMAPP. Baseline covariate in primary model.
[mmapp_total]	Total MMAPP score (computed)	Continuous	Sum of mmapp_1–25 (range 0–75; higher = greater distress)	Dual primary outcome; elevated distress subsample defined as score >10 at baseline

IV. Secondary outcomes

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
F. Secondary outcome — Behavioral Problems (DBIS)				
dbis_1 – dbis_8	8 items measuring disruptive behavior in past two weeks (e.g., 'have you been naughty, stubborn, or acted disrespectfully?')	Ordinal (per item)	0 Never; 1 Sometimes; 2 Often; 3 Very often; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at all timepoints.
[dbis_total]	Total DBIS score (computed)	Continuous	Sum of dbis_1–8 (range 0–24; higher = more problems)	Secondary outcome in primary effectiveness model
H. Secondary outcome — Health-Related Quality of Life (KIDSCREEN-10)				
kidscreen_1 – kidscreen_10	10 items on HRQoL over past two weeks (e.g., 'have you felt fit and well?')	Ordinal (per item)	0 Never; 1 Seldom; 2 Quite often; 3 Often; 4 Always; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at baseline and endline only.
[kidscreen_total]	Total KIDSCREEN-10 score (computed)	Continuous	Sum of kidscreen_1–10 (range 0–40)	Secondary outcome; endline compared between arms adjusted for baseline
I. Secondary outcome — Mental Health Risk Factors (IDEA-RS)				
Idea_rs_run_away_home_1	Ever run away from home	Binary	0 No; 1 Yes; 88 PNTA; 99 DK	IDEA-RS component item; secondary outcome
Idea_rs_run_away_home_2	Ever separated from parents and had to stay elsewhere	Binary	0 No; 1 Yes; 88 PNTA; 99 DK	IDEA-RS component item; secondary outcome
Idea_rs_getting_into_fights_1	In the last year, got into a fight where someone was hurt	Binary	0 No; 1 Yes; 88 PNTA; 99 DK	IDEA-RS component item; secondary outcome

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
Idea_rs_relationship_2	How often talks to caregiver about thoughts and feelings	Ordinal	1 Never; 2 Sometimes; 3 Often; 88 PNTA; 99 DK	IDEA-RS component item; secondary outcome
idea_rs_relationship_3	How often talks to another household member about thoughts and feelings	Ordinal	1 Never; 2 Sometimes; 3 Often; 88 PNTA; 99 DK	IDEA-RS component item; secondary outcome
idea_rs_relationship_4	How often people in home talk about thoughts and feelings	Ordinal	1 Never; 2 Sometimes; 3 Often; 88 PNTA; 99 DK	IDEA-RS component item; secondary outcome
idea_rs_gender	Use child_gender variable	Binary	2 Female is considered risk group	IDEA-RS component item; secondary outcome
idea_rs_minority	Use child_nationality variable	Binary	All groups except Ugandan are considered risk group	IDEA-RS component item; secondary outcome
idea_rs_academic	Out of school	Binary	Not in school at follow-ups	IDEA-RS component item; secondary outcome
idea_rs_druguse	substance_1	Binary	Any substance use	IDEA-RS component item; secondary outcome
Idea_rs_maltreatment	bpq_harsh_parenting	Binary	Any harsh parenting	IDEA-RS component item; secondary outcome
Idea_rs_peer	friendship_1	Binary	No close friends is risk factor	IDEA-RS component item; secondary outcome
Idea_rs_total	Number of IDEA-RS domains	Continuous	Range of 0 to 11	IDEA-RS total score; secondary outcome
J. Secondary outcome — Stigma (Everyday Discrimination Scale, EDS)				

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
eds_1 – eds_8	8 items on frequency of discrimination experiences in daily life (e.g., 'how often are you treated in a way that is not as good as other people?')	Ordinal (per item)	0 Never; 1 <once/year; 2 Few times/year; 3 Few times/month; 4 ≥once/week; 5 Almost daily; 88 PNTA; 99 DK	Secondary outcome (stigma). Sum score used. Assessed at baseline and endline.
[eds_total]	Total EDS score (computed)	Continuous	Sum of eds_1–8 (range 0–40; higher = more discrimination)	Secondary outcome; endline compared between arms adjusted for baseline
eds_9	Open-ended: why does person think they are treated this way	Free text	Free text (conditional on any eds_1–8 > 0)	Contextual/descriptive; not included in main analysis
eds_10	Who treats the child this way	Multi-select nominal	1 Teachers; 2 Classmates; 3 Other school children; 4 Parents/caregivers; 5 Other	Descriptive / contextual for stigma secondary outcome
eds_11	Distress impact of discrimination (0–10)	Continuous	0 Least distressed – 10 Most distressed; 88 PNTA; 99 DK	Descriptive severity indicator for stigma
K. Secondary outcome — Hope (Children's Hope Scale, CHS)				
hope_1 – hope_6	6 items assessing hope (agency and pathways thinking; e.g., 'I think I am doing pretty well')	Ordinal (per item)	0 None of the time; 1 A little; 2 Some; 3 A lot; 4 Most; 5 All the time; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at all timepoints.
[hope_total]	Total CHS score (computed)	Continuous	Sum of hope_1–6 (range 0–30;	Secondary outcome; endline compared between arms

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
			higher = greater hope)	adjusted for baseline
L. Secondary outcome — Social Connectedness (Social Connectedness Scale, SCS)				
connectedness_1 – connectedness_8	8 items on social connectedness (reverse scored; e.g., 'I feel disconnected from the world around me')	Ordinal (per item)	0 Strongly disagree; 1 Disagree; 2 Slightly disagree; 3 Agree; 4 Strongly agree; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at baseline and endline.
[connectedness_total]	Total SCS score (computed)	Continuous	Sum of connectedness_1–8 (range 0–32; higher = less connected [reverse scale])	Secondary outcome; endline compared between arms adjusted for baseline
M. Secondary outcome — Substance Use				
substance_1	Frequency of alcohol consumption in past 3 months	Ordinal	0 None; 1 Once/twice; 2 1–3×/month; 3 1–4×/week; 4 Daily; 88 PNTA; 99 DK	Secondary outcome (substance use). Assessed at baseline and endline.
substance_2	Whether friends consume alcohol	Binary	0 No; 1 Yes; 88 PNTA; 99 DK	Secondary outcome (contextual substance use)
substance_3	Whether friends got drunk in past 3 months	Binary	0 No; 1 Yes; 88 PNTA; 99 DK	Secondary outcome (peer substance use)
N. Secondary outcome — Friendship				
friendship_1	Number of close friends	Continuous	Free integer	Secondary outcome — social functioning
friendship_2	Frequency of socializing with close friends outside school	Ordinal	0 Never; 1 Rarely; 2 Sometimes; 3 Often; 4 Very often; 88 PNTA; 99 DK	Secondary outcome — social functioning

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
friendship_3	Number of close friends who are refugees	Continuous	Free integer	Secondary outcome: refugee peer network (descriptive)
O. Secondary outcome (caregiver) — Psychosocial Wellbeing (WEMWBS)				
WEMWBS_1 – WEMWBS_14	14 items on mental wellbeing over the past 2 weeks (e.g., 'I've been feeling optimistic about the future')	Ordinal (per item)	0 Never; 1 Rarely; 2 Sometimes; 3 Often; 4 Always; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at baseline and endline.
[wemwbs_total]	Total WEMWBS score (computed)	Continuous	Sum of WEMWBS_1–14 (range 0–56; higher = greater wellbeing)	Secondary outcome; endline compared between arms adjusted for baseline
P. Secondary outcome (caregiver) — Psychological Distress (K6)				
KESSLER_1 – KESSLER_6	6 items on psychological distress in the past 2 weeks (e.g., 'how often have you felt nervous')	Ordinal (per item)	0 None of the time; 1 A little of the time; 2 Some of the time; 3 Most of the time; 4 All the time; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at baseline and endline.
[kessler_total]	Total K6 score (computed)	Continuous	Sum of KESSLER_1–6 (range 0–24; higher = greater distress)	Secondary outcome; endline compared between arms adjusted for baseline
Q. Secondary outcome (caregiver) — Parenting (BPQ)				
BPQ_1 – BPQ_23[1] (reverse coded items: BPQ_5, BPQ_15, BPQ_17,	23 items assessing positive parenting (e.g.,	Ordinal (per item)	1 Never; 2 Rarely; 3 Sometimes; 4 Often; 5 Always; 88 PNTA; 99 DK	Secondary outcome. Sum score used. Assessed at

[1] In the original publication of the BPQ, the scale contained 24 items. Due to the authors reporting a low factor loading for item 23, we did not include it in the study.

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
BPQ_18, BPQ_19, BPQ_20)	'when I see my child, I feel happy')			baseline and endline.
[bpq_warmth_parenting]	Total score of BPQ items 1-4, 6-14, 16, 21-23 (computed)	Continuous	Sum of warm parenting items (range 17–85; higher = warmer parenting)	Secondary outcome; endline compared between arms adjusted for baseline
[bpq_harsh_parenting]	Total score of BPQ items 5, 15, 17, 18, 19, 20	Continuous	Sum of harsh parenting items (range 6-30; higher = harsher parenting)	Secondary outcome; endline compared between arms adjusted for baseline

V. Research assistant masking

Variable name	Variable description	Variable type	Variable coding / values	Role in SAP
S. Research assistant masking				
Masking1	RA belief: was child selected for MHPSS programme?	Binary	0 No; 1 Yes	RA masking check — reported in manuscript
Masking2	RA certainty about masking belief	Ordinal	0 Not at all; 1 A bit; 2 Very	RA masking check — reported in manuscript

2. Characteristics of the sample at baseline (Section 25b)

	Overall sample (n=XX)	MHPSS (n=XX)	Waitlist Control (n=XX)
Location, n (%)			
Bukere P/S.	n (%)	n (%)	n (%)
Bukinda P/S	n (%)	n (%)	n (%)
Burungamo P/S	n (%)	n (%)	n (%)
Itambabiniga p/s	n (%)	n (%)	n (%)
Kabazana P/S	n (%)	n (%)	n (%)
Kakoni P/S	n (%)	n (%)	n (%)
Karintuma P/S	n (%)	n (%)	n (%)
Karuhinda P/S	n (%)	n (%)	n (%)
Kaseeta P/S	n (%)	n (%)	n (%)
Kigede Muslim P/S	n (%)	n (%)	n (%)
Kisaaru P/S	n (%)	n (%)	n (%)
Kisabya P/S	n (%)	n (%)	n (%)
Kisagazi P/S	n (%)	n (%)	n (%)
Kyamagabo P/S	n (%)	n (%)	n (%)
Mukondo p/s	n (%)	n (%)	n (%)
Nakivale P/S	n (%)	n (%)	n (%)
Nzozi P/S	n (%)	n (%)	n (%)
Wairagaza P/S	n (%)	n (%)	n (%)
Child age, M (SD)	M (SD)	M (SD)	M (SD)
Child sex, n (%)			
Female	n (%)	n (%)	n (%)
Male	n (%)	n (%)	n (%)
Child nationality, n (%)			
Ugandan	n (%)	n (%)	n (%)
Congolese	n (%)	n (%)	n (%)
South Sudanese	n (%)	n (%)	n (%)
Burundian	n (%)	n (%)	n (%)
Rwandese	n (%)	n (%)	n (%)
Other (Specify)	n (%)	n (%)	n (%)
Child language, n (%)			
Kinyabwisha	n (%)	n (%)	n (%)
Kiswahili	n (%)	n (%)	n (%)
Runyankole	n (%)	n (%)	n (%)
Runyoro-Rutooro	n (%)	n (%)	n (%)

Other	n (%)	n (%)	n (%)
Caregiver age, M (SD)	M (SD)	M (SD)	M (SD)
Caregiver sex, n (%)			
Female	n (%)	n (%)	n (%)
Male	n (%)	n (%)	n (%)
Caregiver nationality, n (%)			
Ugandan	n (%)	n (%)	n (%)
Congolese	n (%)	n (%)	n (%)
South Sudanese	n (%)	n (%)	n (%)
Burundian	n (%)	n (%)	n (%)
Rwandese	n (%)	n (%)	n (%)
Other (Specify)	n (%)	n (%)	n (%)
Caregiver language, n (%)			
Kinyabwisha	n (%)	n (%)	n (%)
Kiswahili	n (%)	n (%)	n (%)
Runyankole	n (%)	n (%)	n (%)
Runyoro-Rutooro	n (%)	n (%)	n (%)
Other	n (%)	n (%)	n (%)
Child sex, n (%)			
Female	n (%)	n (%)	n (%)
Male	n (%)	n (%)	n (%)
Caregiver Relationship to child, n (%)			
Father	n (%)	n (%)	n (%)
Mother	n (%)	n (%)	n (%)
Sibling	n (%)	n (%)	n (%)
Other	n (%)	n (%)	n (%)

3. Table Shell for Outcome Models (Section 27a)

	MHPSS	WLC	Difference in change from baseline to endline	Cohen's d
Primary Outcomes				
Psychological Distress				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M(SD)		
Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d
Wellbeing				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M (SD)		
Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d
Secondary Outcomes (adolescent level)				
Behavioral problems				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M (SD)		
Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d
Health-related quality of life				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M (SD)		
Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d
Stigma				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M (SD)		
Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d
Hope				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M (SD)		
Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d
Social Connectedness				
Baseline	M (SD)	M (SD)		
Midline	M(SD)	M (SD)		
Endline	M (SD)	M (SD)		

Change from baseline to endline	B (95% CI)	B (95% CI)	B (95% CI)	d