Hope Groups: Parenting and Mental Health Among Refugees in the Middle East

Statistical Analysis Plan

June 8, 2026

V1.1

NCT Number: NCT07506421

Protocol ID: 760866

1. Administrative Information

Title of Trial	Hope Groups: A Randomised Controlled Trial Of Psychosocial And Parenting Support Groups For Palestinian Caregivers Affected By War In The Middle East
Trial registration	Open Science Framework Preregistration, January 26, 2025. Registration DOI: https://osf.io/ewaq4 Carbon-copy registered on clinicaltrials.gov. NCT Number: NCT07506421
SAP Version	*This Version 1.0 Statistical Analysis Plan (SAP) for the final trial analysis was written on May 24, 2026, then finalised and posted on June 5, 2026. Version 1.1 was written and posted on June 8, 2026.*

1.1 Roles and Responsibilities

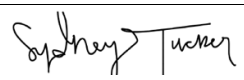
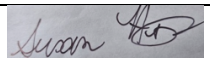
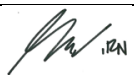
This SAP was produced by members of the trial team as described in the table:

Name	Affiliation(s)	Role
Sydney Tucker, MPH	University of Oxford, Imperial College London	Co - Principal Investigator
Susan Hillis, PhD	Global Reference Group for Children Affected by Crisis	Co - Principal Investigator
Jamie Lachman, DPhil	University of Oxford	Co - Principal Investigator
Oliver Ratmann, PhD	Imperial College London	Statistician and Trial Advisor
G.J. Melendez-Torres, DPhil	University of Exeter	Statistician and Trial Advisor

1.2 SAP revisions

SAP revisions				
SAP revision number	Date of revision	Timing of SAP revision in relation to analysis	Details of revision	Justification for revision
1.1	June 8, 2026	Prior to analysis	Our primary and secondary outcomes are unchanged. In addition to these key prespecified composite outcomes, we have prespecified that we will analyse outcomes separately by subscale as tertiary outcomes.	Analysing primary and secondary outcomes separately by subscale as tertiary outcomes will provide greater detail on intervention effectiveness. This addition is prior to any analysis.

2. Signatures

Name	Signature	Role
Sydney Tucker, MPH		Principal Investigator
Susan Hillis, PhD		Co-Principal Investigator
Oliver Ratmann, PhD		Statistician and Trial Advisor
G.J. Melendez-Torres		Statistician and Trial Advisor
Jamie Lachman, DPhil		Co-Principal Investigator

1. Introduction

1.1 Background

Over 2 billion people live in countries affected by conflict, fragility, or violence globally, and caregivers and children are acutely at-risk. Therefore, there is an urgent need for effective interventions to strengthen families in war settings. In 2023-2024, we conducted a randomised controlled trial (RCT) testing Hope Groups among Ukrainians affected by war.[1] Hope Groups is a psychosocial, mental health, and parenting support group programme delivered by lay-trained facilitators. Among Ukrainians, Hope Groups were found to be highly effective in improving caregiver mental health, strengthening parenting practices, reducing violence against children, and improving child well-being.[1] Based on these results, we are adapting, implementing, and evaluating the Hope Groups intervention in a new setting, among Palestinians experiencing multi-generational displacement in the Middle East.

1.2 Trial Objectives and Hypotheses

The primary objective is to evaluate the causal effects of 10-session Hope Groups among Palestinian parents and caregivers experiencing protracted displacement in Jordan, compared to a wait-list control group, on primary outcomes of caregiver mental health and violence against children post-intervention. The study team hypothesises that the intervention will lead to improvements across both primary outcomes at post-intervention.

The secondary objectives are 1) to evauate the post-intervention effectiveness on secondary outcomes of parenting practices and child behavioural issues and 2) to evaluate the long-term effectiveness of the intervention across primary and secondary outcomes at six- and twelve-months post-intervention.

Additional exploratory objectives include assessing intervention effects on tertiary exploratory outcomes of caregiver well-being, including hopefulness and post-traumatic growth, and physical health, including blood pressure and heart rate.

2. Study methods

2.1 Trial design

This study is a two-armed individually randomised controlled trial utilising a staircase design.

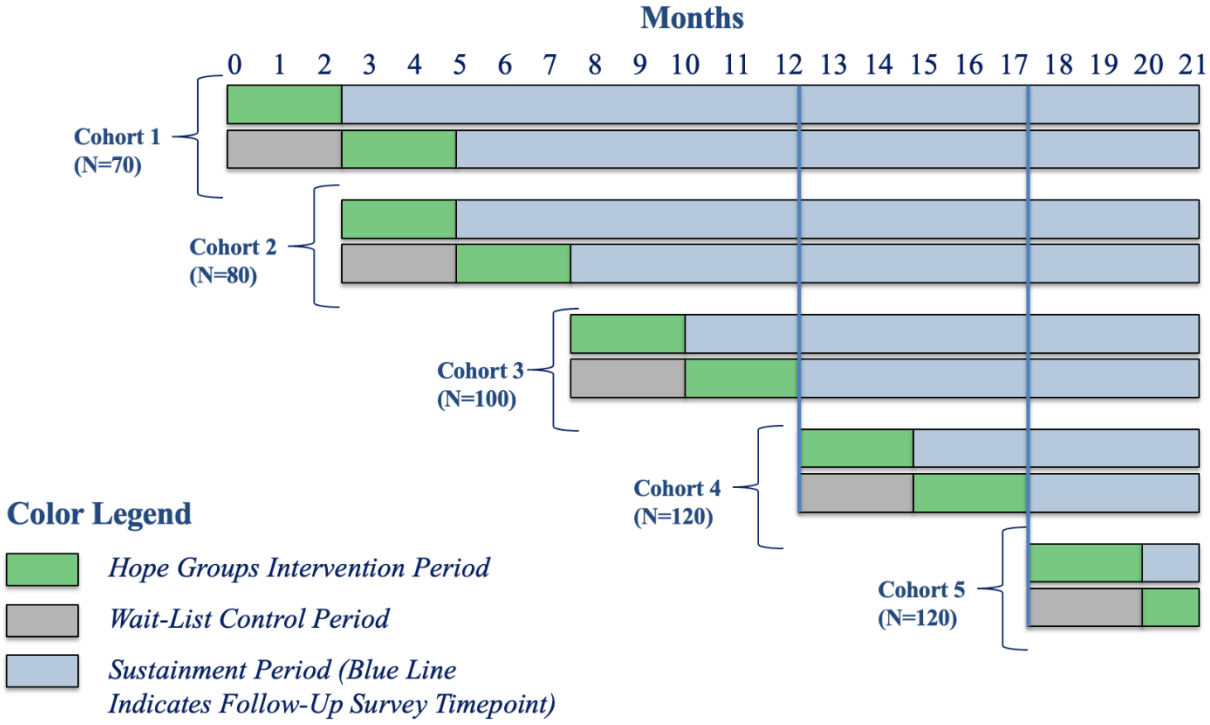
Staircase trials are an innovative adaptation of stepped wedge trials which reduce participant measurement burden [2]. Consistent with stepped wedge trials, the intervention is randomly rolled out to participants across staggered timepoints (also sometimes referred to as “cohorts”). Unlike traditional stepped wedge trials, all participants are not required to complete outcome assessments at each timepoint [2], which drastically increases the feasibility of the design amidst war and displacement contexts.

A staircase RCT was selected to estimate immediate post-intervention and long-term intervention effects. Consistent with a staircase design, participants will be randomly selected

into one of five cohorts (Figure 1). After random selection into one of these cohorts, participants are subsequently randomised with a 1:1 ratio to the intervention arm, which involves participating in *Hope Groups* immediately, or a wait-list control arm which involves no treatment until the cohort's intervention arm reaches trial endline. (Figure 1).

The decision to conduct the trial over five cohorts was made to 1) respect staff capacity relating to the number of participants who could receive the intervention during a set timeframe; and 2) enable six- and twelve- month follow-up assessments among early cohorts (i.e., cohorts 1-3), who will be compared with untreated participants from later cohorts (i.e., cohorts 4 and 5).

Figure 1. Randomised Roll-Out Timeline and Study Design Overview



2.2 Randomisation

The study's lead analyst (ST) will use the slice_sample function within the dplyr package in R (v1.2.1) to randomly select participants for each cohort. Subsequently, within each cohort, individual randomisation with a 1:1 allocation ratio to the intervention arm or wait-list control arm will be conducted using the treatment_assign() function within the RCT package in R (v1.1.2), stratified by sex, by the blinded analyst (ST).

After random allocation, individuals will be subsequently randomly compiled into intervention delivery groups of 8-12 participants (separated by sex) and randomly assigned to a facilitator of the same sex.

Blinding for participants and study staff is not possible, as both parties need to know who is participating in the Hope Groups; however, analysts will be blinded to participant identity during randomisation and all future analyses.

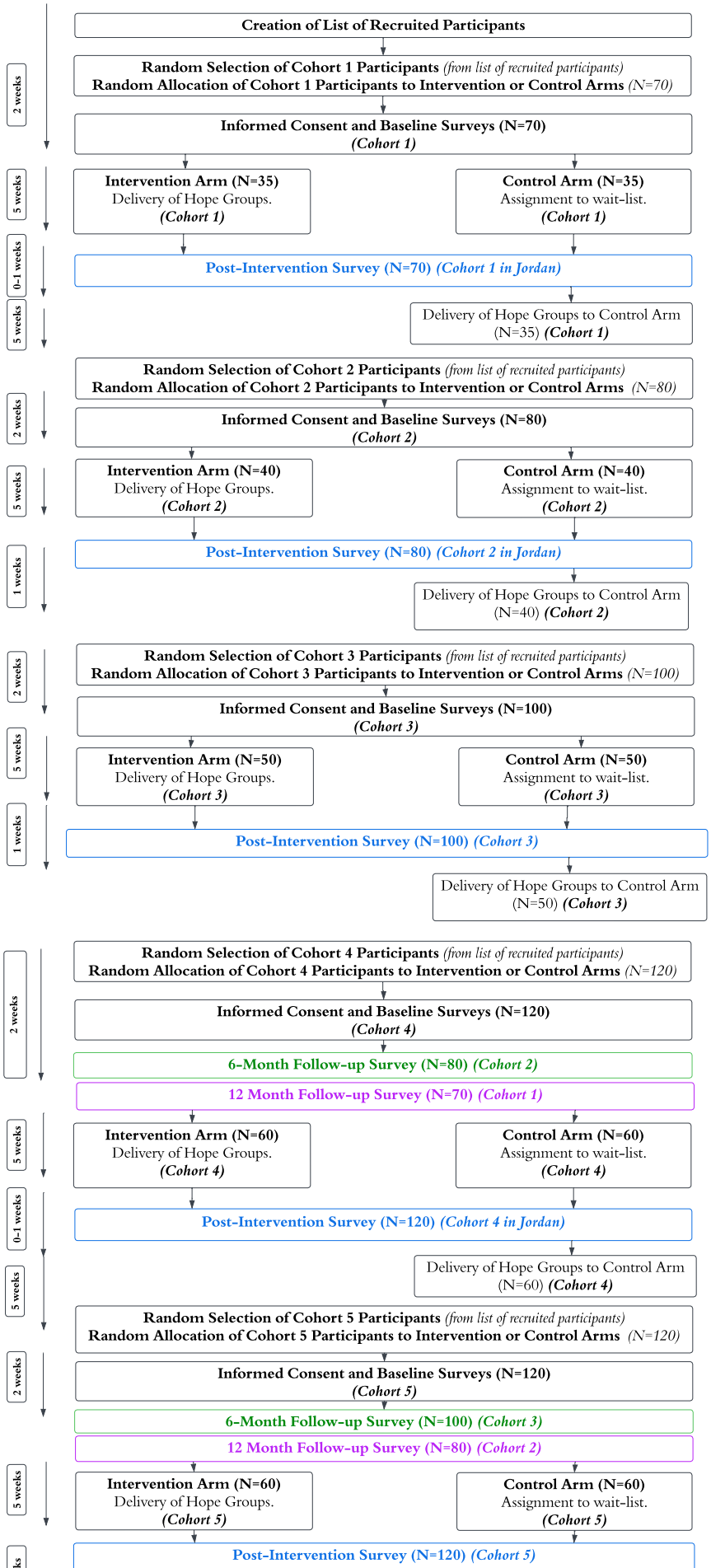
2.3 Sample size

Sample size calculations were initially conducted using the pwr package in R (v 1.3-0). Assuming a two-sided alpha of 0.05, 80% power, 10% loss-to-follow-up, and estimated effect sizes not less than 20% of effects observed in Ukraine, the trial was powered to detect a statistically significant absolute reductions in violence against children and caregiver depression and anxiety symptoms at post-intervention and six- and twelve- month follow-up. Sample size calculations were confirmed using rollout package in R, assuming categorical time effects across each cohort (v 0.1.0).

2.4 Timing of outcome assessments

The trial will conclude after approximately 21 months (Figure 1). Participants will be randomly selected into one of five cohorts. Within cohort, participants will be subsequently randomised to intervention or wait-list control arm (T0). All cohorts complete baseline (T0) and post-intervention surveys (T1) (Figure 2 and 3). Following baseline data collection, the intervention arm will receive the 10-session intervention. One-week after completion of the 10-session intervention, intervention and wait-list control arms will complete a trial endline survey (T1) (Figure 2 and 3). The wait-list control arm will begin the intervention following completion of the endline survey (T1). This procedure will be followed for all cohorts. At the time of randomisation and baseline survey collection for Cohort 4, participants from Cohort 1 and 2 will additionally complete six- and twelve- month long-term follow-up assessments (T2), to be compared with untreated participants in Cohort 4 (T0) (Figure 1 and 2). Similarly, at the time of randomisation and baseline survey collection for Cohort 5, participants from Cohort 2 and 3 will additionally complete six- and twelve- month long-term follow-up assessments (T2), to be compared with untreated participants in Cohort 5 (T0) (Figure 1 and 2).

Figure 2. Trial Flow Chart


Creation of List of Recruited Participants
2 weeks
Random Selection of Cohort 1 Participants (from list of recruited participants)
Random Allocation of Cohort 1 Participants to Intervention or Control Arms (N=70)
Informed Consent and Baseline Surveys (N=70)
(Cohort 1)
5 weeks
Intervention Arm (N=35)
Delivery of Hope Groups.
(Cohort 1)
Control Arm (N=35)
Assignment to wait-list.
(Cohort 1)
0-1 weeks
Post-Intervention Survey (N=70) (Cohort 1 in Jordan)
5 weeks
Delivery of Hope Groups to Control Arm
(N=35) (Cohort 1)
Random Selection of Cohort 2 Participants (from list of recruited participants)
Random Allocation of Cohort 2 Participants to Intervention or Control Arms (N=80)
2 weeks
Informed Consent and Baseline Surveys (N=80)
(Cohort 2)
5 weeks
Intervention Arm (N=40)
Delivery of Hope Groups.
(Cohort 2)
Control Arm (N=40)
Assignment to wait-list.
(Cohort 2)
1 weeks
Post-Intervention Survey (N=80) (Cohort 2 in Jordan)
Delivery of Hope Groups to Control Arm
(N=40) (Cohort 2)
2 weeks
Random Selection of Cohort 3 Participants (from list of recruited participants)
Random Allocation of Cohort 3 Participants to Intervention or Control Arms (N=100)
Informed Consent and Baseline Surveys (N=100)
(Cohort 3)
5 weeks
Intervention Arm (N=50)
Delivery of Hope Groups.
(Cohort 3)
Control Arm (N=50)
Assignment to wait-list.
(Cohort 3)
1 weeks
Post-Intervention Survey (N=100) (Cohort 3)
Delivery of Hope Groups to Control Arm
(N=50) (Cohort 3)
2 weeks
Random Selection of Cohort 4 Participants (from list of recruited participants)
Random Allocation of Cohort 4 Participants to Intervention or Control Arms (N=120)
Informed Consent and Baseline Surveys (N=120)
(Cohort 4)
6-Month Follow-up Survey (N=80) (Cohort 2)
12 Month Follow-up Survey (N=70) (Cohort 1)
5 weeks
Intervention Arm (N=60)
Delivery of Hope Groups.
(Cohort 4)
Control Arm (N=60)
Assignment to wait-list.
(Cohort 4)
0-1 weeks
Post-Intervention Survey (N=120) (Cohort 4 in Jordan)
5 weeks
Delivery of Hope Groups to Control Arm
(N=60) (Cohort 4)
Random Selection of Cohort 5 Participants (from list of recruited participants)
Random Allocation of Cohort 5 Participants to Intervention or Control Arms (N=120)
2 weeks
Informed Consent and Baseline Surveys (N=120)
(Cohort 5)
6-Month Follow-up Survey (N=100) (Cohort 3)
12 Month Follow-up Survey (N=80) (Cohort 2)
5 weeks
Intervention Arm (N=60)
Delivery of Hope Groups.
(Cohort 5)
Control Arm (N=60)
Assignment to wait-list.
(Cohort 5)
Post-Intervention Survey (N=120) (Cohort 5)

3. Statistical Principles

1.1 Uncertainty

Uncertainty intervals will be specified at the 95% level.

1.2 Intervention adherence and protocol deviations

Analysis will be conducted at the individual participant level. Our primary analysis will be intention-to-treat (ITT). If there is significant non-adherence, we will conduct both per-protocol analysis and instrumental variable analysis for exploration, and we will present these results with ITT results.

Protocol deviations, if they occur, are likely to relate to changing needs in a displacement context. Any deviations will be updated on Open Science Framework and in this SAP.

The final trial analysis will be conducted after all cohorts reach the post-intervention timepoint (estimated in July 2026).

4. Trial population

4.1 Eligibility

Participants are recruited through charity partners within the refugee camps we are collaborating with. Inclusion criteria are as follows:

1. Participant is living in one of our partner refugee camps in Amman, Jordan and East Jerusalem, West Bank.

2. Participant is a parent/caregiver for one or more child (of any age).

3. Participant is over the age of 18.

4. Participant has high, medium, or low literacy.
   Note: This is in order to use self-led Hope Group guide and complete surveys. If pilot demonstrates effective results, funding will be sought to create low-literacy versions of Hope Groups guides, using contextualised graphics.

5. Participant is identified by the charity who serves within the shelter as having a demonstrated need for Hope Groups, defined as someone who would benefit from mental health and/or parenting support.
   Note: There are not rigorous requirements for meeting this threshold. The charity has been advised to identify potential participants within the camp who have expressed depression or anxiety, may have a child with behavioural issues, or may have expressed challenges with parenting or specifically talking with their children about war and displacement. Charities have indicated that these issues are unfortunately all very common among parents in the camp, which indicates it is very unlikely that any parent

will be excluded from not meeting this criteria.

6. Participant consents to participate in the study.

4.2 Recruitment

A list of eligible participants will be created via a two-step process. First, the charity operating in Jerash Camp will compile a list of adult parents/caregivers, who will be identified as potentially benefitting from mental health or family strengthening resources. Sanadak, the implementing partner, will call participants to screen for eligibility. During this call, the programme director will communicate more information about the study. Participants will be told the intervention would start either immediately or in several months, depending on random allocation. Eligible participants indicating interest will be kept on the list of participants. For each cohort, participants will be randomly selected from the list and subsequently 1:1 randomised to immediate intervention or wait-list control arm. For each cohort, the programme director will call cohort participants again to invite them to attend an in-person meeting in Jerash Camp, where they will complete informed consent and a baseline survey (T0).

At the in-person meeting, participants will be given a participant information sheet and time to ask any questions prior to completing informed consent. To complete informed consent and baseline surveys (T0), participants will use an online Open Data Kit (ODK) survey form. If participants have a personal phone, they will complete the survey on their device. If not, participants will be given a tablet or phone to complete the survey on. The survey first leads to an informed consent form, then progresses to the baseline survey (taking 25 minutes). If participants are able to read, they will complete the forms independently, to ensure a higher level of anonymity, thereby minimising social desirability bias. If participants need assistance, a facilitator will read the questions to them, then turn the device back to the participant for them to answer the questions privately, aiming to still reduce social desirability bias. After the intervention arm completes the intervention, another in-person data collection event will be held in Jerash Camp for both arms to complete an endline survey (T1). Endline surveys will be completed in ODK, with procedures consistent with T0. Data collection procedures for long-term follow-up survey completion (T2) will be identical with T1.

Data to be reported in the CONSORT diagram are as follows:

- Recruited number of participants
- Participant randomisation to each arm
- Any participant(s) that drop out after randomisation

4.3 Withdrawal

Participants are informed at the time of informed consent that they may withdraw at any time during the study.

5. Statistical analysis

5.1 Baseline characteristics

The following participant baseline characteristics will be collected:

- Age
- Gender
- Marital status
- Education
- Income
- Number of children
- Relationship to children (parent or caregiver)
- Time since displacement

5.2 Primary, Secondary, and Tertiary Outcomes

Primary, secondary, and tertiary outcomes are described in Table 2. Primary outcomes included: 1) caregiver depression and anxiety symptoms measured via Patient Health Questionnaire-4; and 2) violence against children, including caregiver self-report of physical and emotional violence, measured via ICAST-Trial (Table 2). Two primary outcomes were selected as both were considered crucial for strengthening families amidst displacement. Consistent with a pragmatic trial design, effectiveness will not be defined as requiring effects across both primary outcomes; each outcome will be evaluated individually to assess effectiveness. Secondary outcomes included: 1) child behavioural issues, measured via the internalising and externalising subscales from Child and Adolescent Behavioural Inventory (CABI), including items to measure oppositional-defiant disorder and conduct disorder, depression, and irritability; 2) parenting practices, measured via six questions from Alabama Parenting Questionnaire (APQ), as done in the Ukrainian Hope Groups evaluation.[1] This abbreviated APQ scale was selected due to local guidance that the full APQ scale was too lengthy and burdensome for participants. Adaptations required for the local context for primary and secondary outcomes are detailed in Supplementary Materials (SM) Table 1.

To explore a broad range of potential outcomes which could be improved through family strengthening, tertiary exploratory outcomes include 1) caregiver physical health, including blood pressure and heart rate; and 2) well-being, including domains such as quality of life and post-traumatic growth (Table 2). Full details on the survey questions included in tertiary outcomes are included in SM Table 1.

Additionally, as tertiary outcomes, we will analyse 1) violence against children separately by subscale (emotional violence and physical violence); 2) caregiver mental health separately by subscale (depression and anxiety); 3) child well-being separately by subscale (depression,

irritability, and oppositional-defiant disorder and conduct disorder); and 4) parenting practices separately by subscale (parental monitoring, parental involvement, and positive parenting).

All outcomes will be collected at T0 (baseline), T1 (post-intervention endline), and T2 (follow-up), except for blood pressure which will only be collected at T0 and T1 for the first two cohorts (due to limited funding). Outcomes were translated by a Palestinian-Jordanian local, then reviewed by two other Arabic speakers, and discussed in English to assure high-quality translations.

Table 2. Primary, Secondary, and Tertiary Outcomes

Outcome Level	**Domain**	**Outcome**	**Source / Measure Description**	**No. of Items**	**Range** *(response option labels)*
Primary Outcomes					
	Caregiver/ Parent Mental Health	Depression and Anxiety	Patient Health Questionnaire (PHQ) – 4 [3]	4	0–12* *(Not at all, Several days, More than half the days, Nearly every day.*
	Violence Against Children	Physical and Emotional Child Maltreatment	ICAST-Trial – Parent Version - Physical Abuse and Emotional abuse subscales [4]	10	0–30*** *(Multiple times per week, once per week, once in the past TWO weeks, Not at all)*
Secondary Outcomes					
	Parenting Practices	Positive Parenting, Parental Involvement, and Parental Monitoring	Alabama Parenting Questionnaire (APQ) [5]	6	0–7** (*Days in the past week)*
	Child Well-being	Internalising and externalising behaviours	Child and Adolescent Behavioural Inventory (CABI) [6]	17	0–32† (*Not true, Sometimes true, Somewhat true, or Very true)*
Tertiary Outcomes					
	Caregiver Physical Health	Blood pressure	Average of two BP readings	4	Systolic: Continuous; Diastolic: Continuous

		Heart rate	Average of two heart rate readings	2	Continuous
		Caregiver sickness	Direct report of number of days caregiver was sick in past week	1	0–7** (*Days in the past week)*
	Caregiver Well-Being	Post-traumatic growth	Post-Traumatic Growth Inventory [7]	3	0–12 § *(Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree.)*
		Quality of life	WHO Quality of Life Scale (The WHOQOL Group, 1998)	3	0–12‡ *(Not at all, A little, Moderately, Completely.)*
		Suicidal thoughts	PHQ-9 [3]	1	0–3* *(Not at all, Several days, More than half the days, Nearly every day.)*
		Coping with grief and hopeful-ness	Locally developed in Ukraine	2	0–7** (*Days in the past week)*
		Self-care	Locally developed in Ukraine	1	0–7** (*Days in the past week)*
		Hope and purpose in community: Beliefs about improving community	Locally developed in Jordan	2	0–8‡ *(Not at all, A little, Moderately, Completely.)*
		Hope and purpose in family: Belief in ability to care for self and children	Locally developed in Jordan	1	0–4‡ (*Not at all, A little, Moderately, Completely.)*

5.3 Descriptive Statistics

Before undertaking the primary analysis, we will examine the mean, median, range, and distribution of all variables at baseline, endline, and follow up (T0, T1, T2). We will prepare a table of summary statistics including numbers and proportions for categorical variables and mean (including standard deviation) and median (including interquartile range) for all continuous variables at baseline. Patterns of missingness will also be examined for all outcomes at item- and scale- level. Histograms and summary statistics will confirm the choice of regression model.

5.4 Analysis Methods

As is standard in staircase trials[2], the post-intervention treatment effect will be estimated by pooling together the treatment effects across all intervention and control arm post-intervention comparisons across the five cohorts. All analyses will use time-by-time models due to shifting populations at each subsequent follow-up. An intention-to-treat approach with a binary exposure variable indicating if the participant was randomised to the intervention or control group will be used. The primary method of analysis will be a generalised linear mixed-effects model, estimating the difference between intervention and control participants (i) post-intervention (T1), controlling for baseline level of the outcome, cohort, facilitator, and sex. The regression link function will be based on consideration of outcome distributions at baseline, and the appropriate link function will be used for each outcome. Where an ideal link function is not available (e.g., limited range variables where Poisson, ordered logit, and identity links are inappropriate), we will use wild bootstrapping. If a bootstrap method is used, percentile-based 95% confidence intervals will be reported. We will use two-tailed tests for all models with statistical significance thresholds of 0.05.

The six-month follow-up treatment effects will be estimated by pooling together the treatment effects across comparisons of participants completing six-month follow-up assessments with untreated participants completing assessments at the same timepoints. Similarly, the twelve-month follow-up effects will be estimated through pooling together participants completing twelve-month follow-up assessments with untreated participants completing assessments at the same timepoints. Across both six- and twelve- month follow-up analyses, an intention-to-treat approach will be used, with a binary exposure variable indicating if the participant is at a follow-up or untreated timepoint. The same statistical model reported for post-intervention analysis will be used for both six- and twelve- month follow-up analyses: The primary method of analysis will be a generalised linear mixed-effects model, estimating the difference between follow-up participants and wait-list control participants (i), controlling for cohort, facilitator, and sex. Exchangeability across cohorts will additionally be assessed through comparing demographic covariates. If any imbalance is detected, propensity-score weighting to adjust for cohort differences will be utilised in six- and twelve- month follow-up analyses.

Post-intervention analyses will include all randomised participants, across cohorts 1-5 (N=490). Consistent with stepped wedge and staircase trials, six-month follow-up analyses will include all participants completing six-month follow-up surveys with all untreated participants completing baseline surveys at the same timepoint (N=420). Twelve-month follow-up analyses will include all participants completing twelve-month follow-up surveys with all untreated participants completing baseline surveys at the same timepoint (N=390).

Adverse event risk differences were estimated using the Newcombe (Wilson-based hybrid) method.

5.5 Additional Exploratory Analyses

To uncover greater detail on intervention effectiveness, as mentioned, we will analyse as tertiary outcomes 1) violence against children separately by subscale (emotional violence and physical violence); 2) caregiver mental health separately by subscale (depression and anxiety); 3) child

well-being separately by subscale (depression, irritability, and oppositional-defiant disorder and conduct disorder); and 4) parenting practices separately by subscale (parental monitoring, parental involvement, and positive parenting).

To try to increase the interpretability of results, when possible, we will conduct additional analyses to estimate the reduction of participants meeting thresholds for clinical conditions. For example, for caregiver mental health, we will explore the difference in participants with symptoms reaching the threshold for depression or anxiety, in the intervention arm versus the control arm.

Additional moderator and subgroup analyses will be identified and preregistered prior to analysis.

5.6 Missingness

If data missingness in any one outcome model exceeds 10%, multiple imputation and inverse probability weighting will be used to account for missingness.

5.7 Statistical Software

All data cleaning and analysis will be completed in R.

6. Harms

Fatal and life-threatening events (otherwise known as serious adverse events) experienced by participants while participating in the study are unlikely to be caused by participation in this trial. Nonetheless, any adverse events reported during the duration of this study will be reported to ethical review boards who approved this study.

7. Quality Assurance

As a minimum, primary analyses will be performed twice by separate analysts. Dependent on capacity, analysis of secondary outcomes will also be undertaken twice.

Supplementary Materials

Supplementary Materials Table 1. Primary, Secondary, and Tertiary Outcomes with Adaptation Explanations and Survey Question Details

Outcome Level	**Domain**	**Outcome**	**Source / Measure Description**	**No. of Items**	**Range** *(response option labels)*	**Adaptations**
Primary Outcomes						
	Caregiver/ Parent Mental Health	Depression and Anxiety	Patient Health Question-naire (PHQ) – 4 (Carroll et al., 2020)	4	0–12* *(Not at all, Several days, More than half the days, Nearly every day.*	None
	Violence Against Children	Physical and Emotional Child Maltreatme nt	ICAST-Trial – Parent Version - Physical Abuse and Emotional abuse subscales (Meinck et al., 2018)	10	0–30*** *(Multiple times per week, once per week, once in the past TWO weeks, Not at all)*	- Due to confusion in local setting, 1 item was removed from the physical violence subscale ("Shake him/her"). - Due to locally-deemed inappropriateness in a Muslim or extremely low-resource refugee camp setting, 3 items were removed from the emotional violence subscale, including 1) "Threaten to invoke ghosts or evil spirits, or harmful people"; 2) "Without a meal as a punishment"; and 3) "Lock him/her out of the house".
Secondary Outcomes						
	Parenting Practices	Positive Parenting, Parental Involve-ment, and Parental Monitoring	Alabama Parenting Question-naire (APQ) (Shelton et al., 1996)	6	0–7** (*Days in the past week)*	Consistent with the Ukrainian Hope Groups evaluation, 2 items from each of the positive parenting, parental involvement, and parental monitoring subscales were chosen, to reduce participant burden.
	Child Well-being	Internal-ising and external-ising behaviours	Child and Adolescent Behavioural Inventory (CABI) (Cianchetti et al., 2013)	17	0–32† (*Not true, Sometimes true, Somewhat true, or Very true)*	- Oppositional-defiant disorder and conduct disorder subscale: Based on cultural relevance and the low-resource context, 7 items were used, and 3 items from this subscale were dropped, including 1)*"he is domineering and always wants to assert himself", which was viewed as positive by local leaders; 2)"he bothers and intentionally annoys others", which was viewed as less important by local leaders;*

						and 3)"he has committed petty theft", which was viewed as highly common by local leaders. - Depression subscale: 8 items on depression were used, including all the validated items on mood, aboulia, and suicidal thoughts. Due to local input on prioritisation in the local setting, the 1 item on self-esteem and 1 item on guilt were not included. - Irritability subscale: 2 items on irritability were used. To shorten the survey, 2 items from this subscale were dropped, which were deemed by local advisors are highly common in a refugee context, and thus yielding less meaningful information, including 1)*"he has frequent mood changes"; and 2)"he is very irritable"*.
Tertiary Outcomes						
	Caregiver Physical Health	Blood pressure	Average of two BP readings	4	Systolic: Continuous; Diastolic: Continuous	NA
		Heart rate	Average of two heart rate readings	2	Continuous	NA
		Caregiver sickness	Direct report of number of days caregiver was sick in past week	1	0–7** (*Days in the past week)*	NA
	Caregiver Well-Being	Post-traumatic growth	Post-Traumatic Growth Inventory (Cann et al., 2010)	3	0–12 § *(Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree.)*	Due to constraints regarding survey length, 3 of the most-relevant items from the 10-item form were selected by local leaders, including: - *"Hardship has taught me I am stronger than I thought I was"* - *"Hardship has taught me that I can handle difficulties."* - *"Hardship has given me a greater sense of closeness with others."*
		Quality of life	WHO Quality of Life Scale (The WHOQOL	3	0–12‡ *(Not at all, A little, Moderately,*	Due to constraints regarding survey length, 3 questions from spiritual, religious, or personal beliefs subscale, including: - *"How much does spiritual strength help you to live better?"*

			Group, 1998)		*Completely.)*	- *"To what extent do you have inner peace?"* - *"To what extend are you hopeful about your life?"*
		Suicidal thoughts	PHQ-9 (Carroll et al., 2020)	1	0–3* *(Not at all, Several days, More than half the days, Nearly every day.)*	NA. The one item on suicidality from PHQ-9 was used.
		Coping with grief and hopeful-ness	Locally developed in Ukraine	2	0–7** (*Days in the past week)*	Locally developed in Ukraine, and determined to be included in Jordan. Consisting of 2 items, developed to measure constructs related to positive psychology and coping with grief: - *How many days in the past week did you feel hopeful about the future?* - *How many days in the past week were you able to grieve your losses in a healthy way, such as by talking with someone you trust, finding ways to honour who or what you lost, or finding meaning in your new reality?*
		Self-care	Locally developed in Ukraine	1	0–7** (*Days in the past week)*	Locally developed in Ukraine, and determined to be included in Jordan, consisting of: - *How many days in the past week did you take actions to care for yourself, such as relaxation techniques, talking to someone you trust about your feelings, establishing a routine, or eating and sleeping well?*
		Hope and purpose in community: Beliefs about improving community	Locally developed in Jordan	2	0–8‡ *(Not at all, A little, Moderately, Completely.)*	Locally developed in Jordan, including the following questions: - *"Do you believe things in your community can get better?"* - *"Do you believe you can be a part of making things in your community better?"*
		Hope and purpose in family: Belief in ability to care for self and children	Locally developed in Jordan	1	0–4‡ (*Not at all, A little, Moderately, Completely.)*	Locally developed in Jordan, including the following questions: - *"Do you believe you have the skills to take care of yourself and your children, regardless of whether circumstances in your community get better or not?"*

References

1. Tucker, S. *et al.* Building Hope in Crisis: A Multi-Country Cluster Randomised Controlled Trial Evaluating a Psychosocial, Parenting, and Violence Prevention Intervention During War in Ukraine. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.5228739 (2026).
2. Grantham, K. L., Forbes, A. B., Hooper, R. & Kasza, J. The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Stat. Methods Med. Res.* **33**, 24–41 (2024).
3. Carroll, H. A. *et al.* Establishing reliability and validity for mental health screening instruments in resource-constrained settings: Systematic review of the PHQ-9 and key recommendations. *Psychiatry Res.* **291**, 113236 (2020).
4. Meinck, F. *et al.* 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 Negl.* **82**, 45–58 (2018).
5. Shelton, K. K., Frick, P. J. & Wootton, J. Assessment of parenting practices in families of elementary school-age children. *J. Clin. Child Psychol.* **25**, 317–329 (1996).
6. Cianchetti, C. *et al.* Child and Adolescent Behavior Inventory (CABI): A New Instrument for Epidemiological Studies and Pre-Clinical Evaluation. *Clin. Pract. Epidemiol. Ment. Health CP EMH* **9**, 51–61 (2013).
7. Cann, A. *et al.* A short form of the Posttraumatic Growth Inventory. *Anxiety Stress Coping* **23**, 127–137 (2010).
8. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. *Psychol. Med.* **28**, 551–558 (1998).