ParentText in Crisis: A Randomised Controlled Trial of Chatbot-Delivered Mental Health and Parenting Content for Palestinian Caregivers Affected by War, Crisis, and Protracted Displacement

Interim Analysis Statistical Analysis Plan

NCT Number: NCT07494786

August 17, 2026

V1.0

1. Administrative Information

Title of Trial	ParentText For Crisis: A Randomised Controlled Trial of Chatbot-Delivered Parenting Content for Palestinian Caregivers in Jordan
Trial registration	Open Science Framework, December 5, 2025. Registration DOI: https://osf.io/hvcje/overview
SAP Version	*This Statistical Analysis Plan (SAP) was written on August 5, 2026; Version 1.0.*

1.1 Roles and Responsibilities

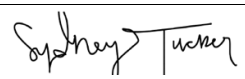
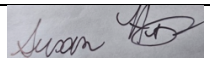
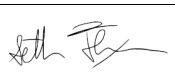
This SAP was produced by members of the trial team as described in the table:

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
Seth Flaxman	University of Oxford	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

2. Signatures

Name	Signature	Role
Sydney Tucker, MPH		Principal Investigator
Susan Hillis, PhD		Co-Principal Investigator
Jamie Lachman, DPhil		Co-Principal Investigator
Oliver Ratmann, PhD		Statistician and Trial Advisor
Seth Flaxman, PhD		Statistician and Trial Advisor

1. Introduction

1.1 Background

Over two-thirds of the world's children live in conflict-affected countries. (*Peace Research Institute Oslo*, 2023). In-person parenting programmes can improve caregiver mental health and reduce violence against children even amidst humanitarian settings (Backhaus et al., 2025); however, in-person programming can be costly and challenging to implement at-scale to the massive number of families in need. Despite the potential scalability of digitally delivered parenting programmes, sparse evidence exists on whether digitally delivered parenting programmes can strengthen parents and protect children amidst conflict. Therefore, evaluating digitally delivered parenting programmes in conflict settings is an important research priority.

To create a low-cost, highly cost-effective digital intervention, Parenting for Lifelong Health, Oxford University, Sanadak, World Without Orphans, and Anar for Empowerment and Psychosocial Support partnered to create "ParentText for Crisis" for Palestinians living in Jerash Camp. ParentText for Crisis is a rule-based chatbot developed with and for Palestinian caregivers. The intervention was adapted from Parenting for Lifelong Health's ParentText chatbot delivery system (Ambrosio et al., 2024) and abbreviated content from an in-person mental health and parenting programme called Hope Groups (Tucker et al., 2026; Tucker et al., 2025), resulting in a 9-session ParentText for Crisis programme. The chatbot uses text, audio, and video to deliver abbreviated Hope Groups content via nine pre-programmed messaging sessions, which include encouraging tips and activities for caregivers to use at home. All nine sessions provide content to 1) help parents cope with trauma, anxiety, and depression; 2) improve parenting practices to keep children healthily developing even amidst crisis and protect children from further trauma, with a high potential for scale given its low delivery cost in comparison to the in-person programmes (Tucker et al., 2025).

This pilot study aims to evaluate the effectiveness of "ParentText for Crisis" among Palestinians displaced in Jerash Camp, Jordan.

1.2 Trial Objectives and Hypotheses

The primary objective is to evaluate the intervention effects of ParentText for Crisis +30 days post-onboarding among Palestinian parents and caregivers experiencing protracted displacement in Jordan, compared to a wait-list control group, on primary outcomes of caregiver depression and playful parenting.

The secondary objectives are to evaluate the intervention effectiveness +30 days post-onboarding on secondary outcomes of violence against children, parental monitoring, and caregiver hopefulness about the future.

2. Study methods

2.1 Trial design

This study is a two-armed parallel-group randomised controlled trial. Participants are individually randomised to intervention, which consists of participating in ParentText for Crisis immediately, or a wait-list control arm, who will receive the intervention after trial endline.

2.2 Randomisation

Individual randomisation to intervention or wait-list control arms will be conducted using the treatment_assign() function in the RCT package in R (v1.1.2) by a blinded analyst (ST). Blinding for participants and study staff is not possible, as both parties need to know who is participating in the intervention; however, analysts will be blinded to participant identity during randomisation and all future analyses.

2.3 Sample size

This research is an exploratory trial to inform a future larger trial. The sample size was set by pilot funding constraints. Initial funding permitted the inclusion of N=240 participants. With a sample of N=240, and assuming 80% power, the minimum detectable effect would be d=0.36. Additional funding was secured to enable the inclusion of an additional N=240 participants in a second cohort, after an interim analysis following the completion of cohort one, which will inform whether the intervention should remain identical, if demonstrating promise, or if the second cohort will test an enhanced intervention. With a sample of N=480, and assuming 80% power, the minimum detectable effect would be d=0.26.

2.4 Timing of outcome assessments

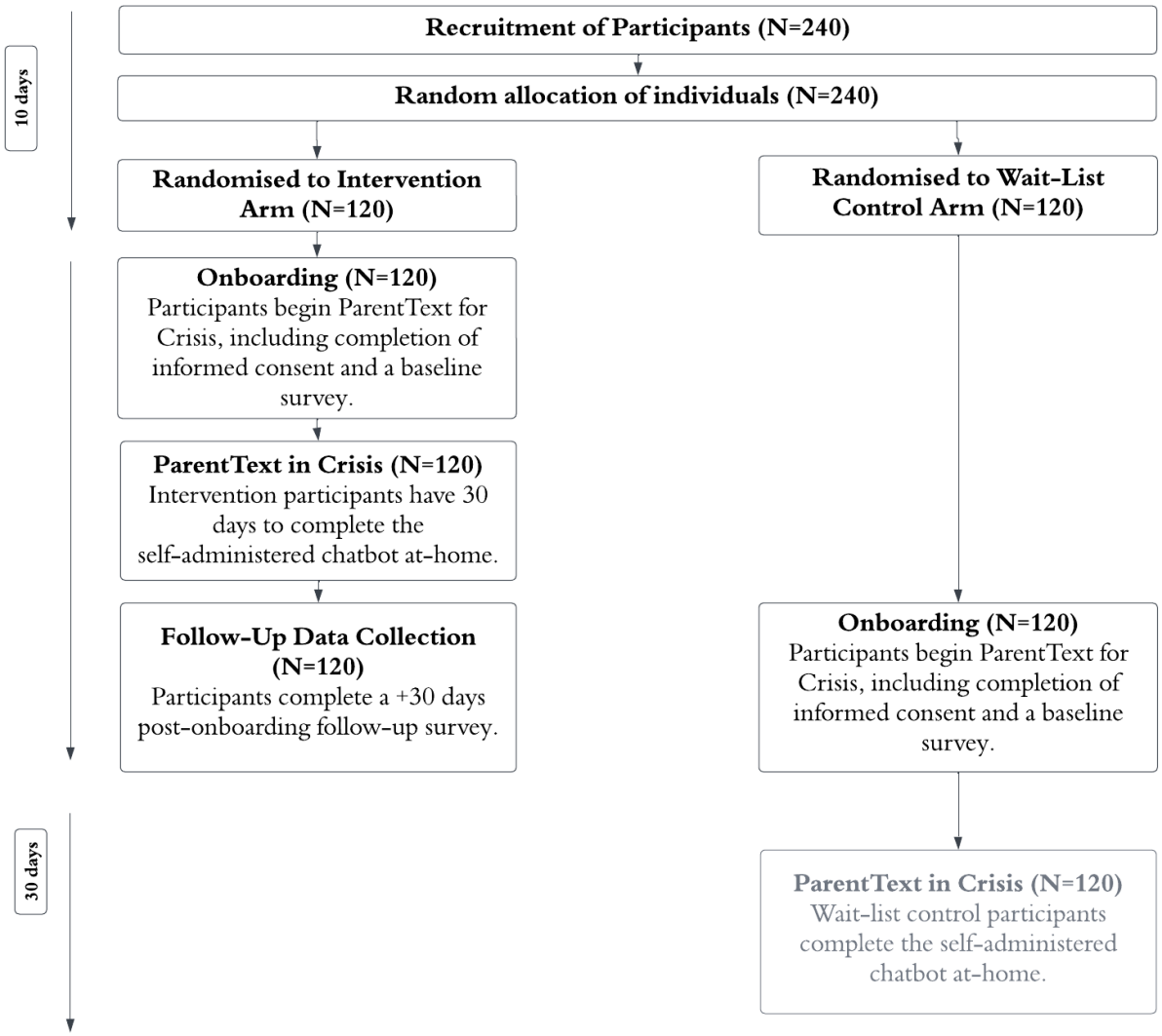
The study will include 480 Palestinian participants 1:1 randomised to intervention arm, who join ParentText for Crisis immediately, or a waitlist control arm. The study rolls out over two cohorts, conducted as two back-to-back RCTs. The first cohort includes N=240 Palestinian participants, and the second cohort includes another N=240 Palestinian participants (Figure 10, A and B). **This statistical analysis plan (SAP) is for the interim analysis, occurring after the first cohort of N=240 participants.**

Across both cohorts, the intervention arm completes a baseline survey at T0, when they onboard into the chatbot intervention, then a follow-up survey at T1, +30 days after intervention onboarding. The control arm takes a baseline survey at T1. Thus, intervention effectiveness is estimated via comparing the treated intervention arm and untreated control arm at T1. A single assessment timepoint was used to maximise study feasibility and reduce implementation costs with a pilot budget in a crisis setting, where requiring control participants to take baseline surveys at the time of randomisation adds substantial costs and logistical complexities. The limitation of this approach is the inability to assess randomisation balance at baseline across outcome measures. Thus, balance achieved from randomisation can only be assessed by comparing non-time-varying covariates between arms, such as age, sex, and number of children.

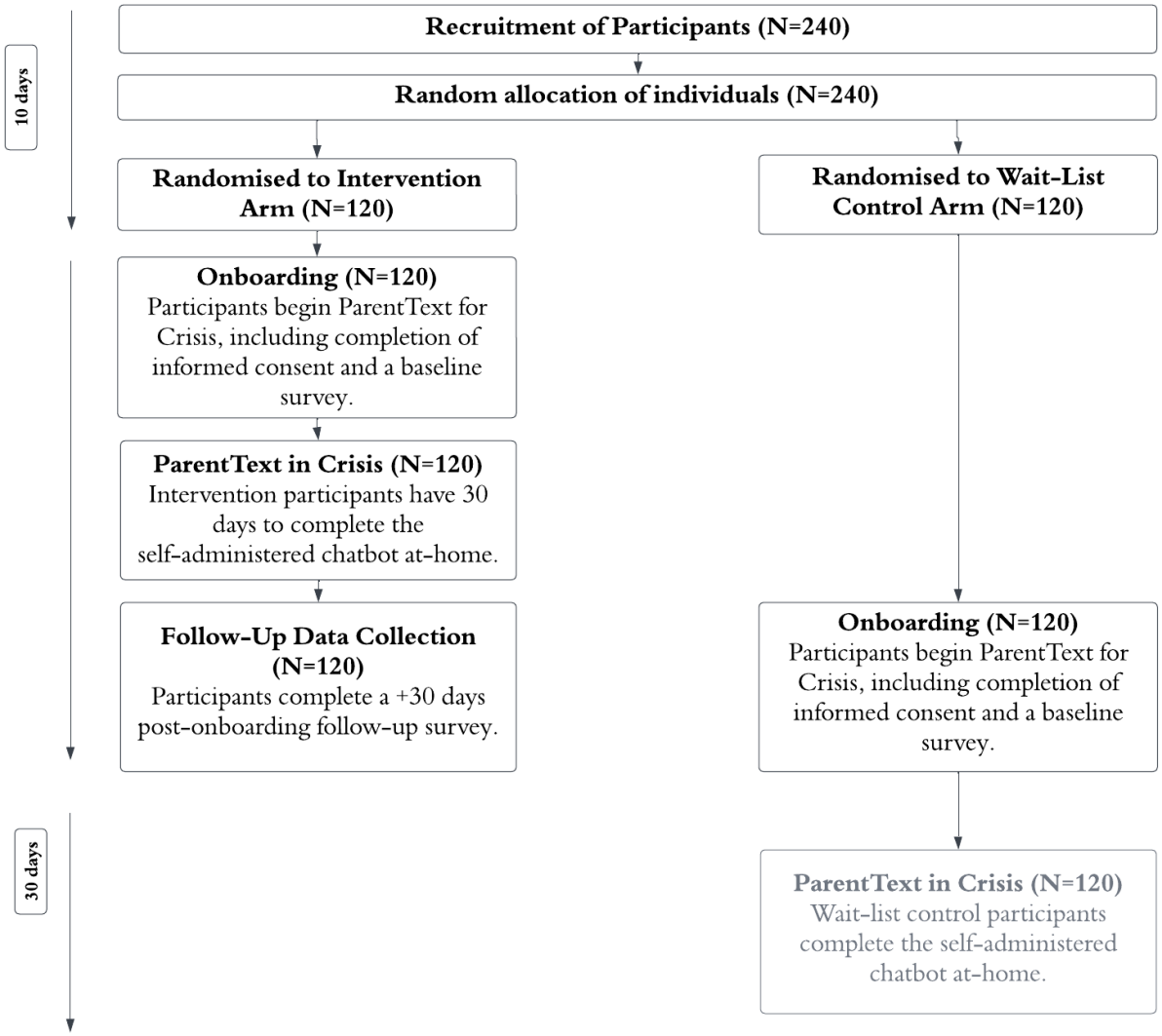
This SAP is for an interim analysis plan after the conclusion of the first cohort, prior to the start of the second cohort (Figure 10, A).

Figure 10. Randomised Controlled Trial Flowchart

A. Cohort 1



B. Cohort 2



2.5 Timing of interim analysis and final analysis

The trial interim analysis will be conducted for primary outcomes after the completion of cohort 1, after the intervention arm reaches +30 days post-intervention (T1). Results from the interim analysis may inform intervention delivery in cohort 2, following an adaptive trial approach.

The final analysis will be conducted after the completion of cohort 2. A final analysis SAP will be registered prior to the analysis including cohort 2. The current statistical analysis plan is for the interim analysis of cohort 1 only.

3. Statistical Principles

3.1 Uncertainty

Uncertainty intervals will be specified at the 95% level.

3.2 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.

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 Jerash Camp.
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 access to a phone with WhatsApp.
5. 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. Participants on the list will be 1:1 randomised to intervention or control arms.

After randomisation, intervention arm participants will be invited to an in-person meeting where they will complete informed consent and intervention onboarding. During this in-person session, participants will be given an overview of the chatbot programme and time to ask questions. They will also receive the trigger word to text to the chatbot's number on WhatsApp to begin the intervention. Participants will send the trigger word during the in-person session, and the chatbot will initiate the process of onboarding, providing an information sheet and terms of use, and informed consent. After informed consent, the chatbot will continue to the baseline survey questions, then initiate session one. Sanadak staff and volunteers will be present at the in-person event to help in the event of participants needing reading support.

After +30 days, intervention arm participants will be invited for a second in-person event. At this session, participants will be given a trigger word to send to the chatbot to receive a follow-up survey. The same day, wait-list control arm participants will be invited for their first in-person event, where they will complete informed consent and the survey assessment. After completion of the survey, they will begin session one of the intervention.

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

All recruitment procedures are conducted identically for cohort 1 and 2.

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
- Time since displacement

5.2 Primary and Secondary Outcomes

To maximise feasibility of chatbot data collection, abbreviated measures were essential. This approach aligns with a larger ParentText RCT conducted in South Africa, but even shorter measures were needed in the crisis setting of Jerash Camp, as digital literacy was low. Primary outcomes included caregiver depression and playful parenting, while secondary outcomes included hopefulness about the future, parental monitoring, and violence against children. All primary and secondary outcomes are detailed in Table 1.

During the chatbot-facilitated survey assessments, the three primary outcome questions were asked first. Subsequently, participants were asked if they were willing to complete another five questions; participants who opted in answered the secondary outcome questions. Utilisation of abbreviated outcomes and optional secondary outcomes measures limits the reliability of these measures. However, this light-touch approach was chosen to enable initial preliminary findings to inform a future larger study, amidst a community with limited exposure and potential scepticism towards chatbot-led programmes. A future study will include more robust measures across parental mental health, parenting practices, violence against children, and child well-being.

The interim analysis will assess primary outcomes only.

Table 1. Primary and Secondary Outcomes

Outcome Level	**Outcome**	**Source**	**No. of Items**	**Scale / Range**
Primary	Caregiver Depression	Patient Health Questionnaire - 2 (Kroenke et al., 2003)	2 items (full scale)	0–6* *(Not at all, Several days, More than half the days, Nearly every day)* **Validated scale defines depression as scores >=3.*
Primary	Playful Parenting	Alabama Parenting Questionnaire (APQ) – Parental Involvement (Shelton et al., 1996)	1 item: - *How many days in the past week did you play with your children or help them learn something new?*	0–7** *(Numbers of days in the past week)*
Secondary	Caregiver Hopefulness	Positive Psychology: Hopefulness about the future	1 item: - *How many days in the past week did you feel hopeful about the future?*	0–7** *(Numbers of days in the past week)*
Secondary	Parenting Practices	Alabama Parenting Questionnaire (APQ) – Parental Monitoring and Supervision (Shelton et al., 1996)	2 items on parental monitoring and supervision *(measures used in Papers 1-3)*	0–7** *(Numbers of days in the past week)*
Secondary	Violence Against Children	Behavioural indicators from ICAST-Trial	2 items *(measures used in Paper 1)*	0–7**

		assessing physical and emotional violence (Meinck et al., 2018)		*(Numbers of days in the past week)*

** Survey questions are on the following ordinal scale to indicate occurrence in the past week: Not at all, Several days, More than half the days, Nearly every day.*
*** Survey questions asked for frequency of occurrence in the number of days in the past week. When multiple items are used, responses are averaged together to result in a range still between 0-7.*

5.3 Descriptive Statistics

Before undertaking any analysis, we will examine the mean, median, range, and distribution of all variables. 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 Approach

Treatment effects will be estimated by comparing intervention arm outcomes at +30 days post-onboarding to control outcomes at the same timepoint. The analysis will compare intervention participants +30 days post-onboarding to wait-list control participants prior to beginning the intervention, with a fixed effect on sex. If any covariates are imbalanced between arms, they will be included as additional fixed effects in the model.

The interim analysis will be conducted after the first cohort of N=240 participants, using a Bayesian approach. Three values will be estimated to assess statistical effects of the intervention:

1. The posterior probability of a protective treatment effect, given the current data (N=240 participants), will be estimated along with a 95% credible interval. This will provide insight on the current treatment effect among N=240 participants, by assessing if the chatbot intervention demonstrates a significant effect across primary outcomes indicated above, as measured by quantifying if posterior credible intervals of the treatment effect include the null with a probability of less than 5%.
2. The predictive probability for a protective effect, with an additional second cohort of N=240 participants (total N=480 participants), will also be estimated. This will provide insight into whether the trial would be expected to find statistical significance, if the trial continued into a second cohort, with the intervention in its current form.
3. E-values will be estimated to additionally quantify the current level of evidence on the intervention's effectiveness at the interim analysis timepoint (Ramdas et al., 2023). The e-value is "anytime valid", defined as able to be calculated at the interim analysis and at all subsequent analyses, without the need for corrections, to demonstrate the strength of evidence at each timepoint (Ramdas et al., 2023), creating a way to quantify evidence at a specific timepoint (e.g., after cohort one) and additionally assess how evidence is building over time (e.g., after cohort two). E-values compare the difference observed between the two trial arms against what would be expected under the null hypothesis of no treatment effect, quantifying how strongly the data support a treatment effect over the null. The expected differences are estimated by permutation, repeatedly shuffling the

treatment assignment labels to generate the null distribution. Under the null hypothesis of no effect, the e-value has an expected value of 1, and the probability that it ever reaches a value c is no greater than 1/c by Ville's inequality. A threshold of 20 therefore corresponds to a one-sided false-positive rate of at most 1/20 = 5%. An e-value of 20 or greater for a specific outcome indicates substantial evidence of effectiveness on that outcome, which would correspond to statistical significance. For each primary outcome, an e-value will be calculated to measure the treatment effect's strength of evidence.

Steps on how the three values listed above will be interpreted are explained in section 5.5.

Additionally, the interim analysis will assess the clinical significance of the intervention, to ensure the intervention not only demonstrates promise for statistical significance but also promise for clinically meaningful significance. To investigate the potential of the intervention to yield clinically meaningful results, the chatbot intervention will be compared with the evidence-based in-person programme which the chatbot content was adapted from (Tucker et al., 2025), accounting for differences in cost. The cost-effectiveness of both interventions will be compared, rather than just comparing the magnitude of treatment effects; the reason for this is that the chatbot is likely to produce smaller treatment effects, as it is a very light-touch digital intervention, *but its lower cost will enable delivery to substantially more families than an in-person programme with the same level of financial investment*. Therefore, it is essential to compare not just the magnitude of treatment effects, but to also account for the substantial difference in delivery cost, which impacts the number of families who in the camp who can participate.

1. This assessment will be conducted for both primary outcomes:
    1. (i) Caregiver depression: The cases of caregiver depression averted in the chatbot intervention arm versus the wait-list control arm will be estimated. Based on this estimated risk difference, using costing information from both the chatbot and the in-person intervention, the estimated number of depression cases averted per $100,000 invested in programming will be estimated for both the chatbot intervention and the in-person intervention. These results will reveal which intervention averts the highest number of depression cases with an equivalent level of investment.
    2. (ii) Playful parenting: The difference in days of utilising playful parenting practices will be estimated in the chatbot intervention arm versus the wait-list control arm. Based on this estimated difference, using costing information from both the chatbot the in-person intervention, the estimated number of days of playful parenting per $100,000 invested in programming will be estimated for both the chatbot intervention and the in-person intervention. These results will reveal which intervention yields the highest number of days where children experience play with their parent/caregiver with an equivalent level of investment.

Beyond this quantitative estimate of cost-effectiveness, we will additionally assess programme dropout rates, to monitor whether the majority of participants (>80%) seem able to engage with and complete the programme. If not, we will collaborate with our implementing partner to

brainstorm additional ways to strengthen engagement prior to cohort two.

Steps on how clinical significance will be interpreted is explained in section 5.5.

5.5 Interim Analysis Review

The decision rules used to interpret results from the interim analysis are as follows.

In summary, if the intervention does not show promise for both statistical significance and clinical significance at the interim analysis timepoint, we will collaborate with the local implementing partner to enhance the intervention prior to the start of the second cohort (e.g., adding an additional in-person session within the chatbot intervention).

1. If harmful effects are indicated, we will PAUSE activity, and we will NOT proceed with delivering the intervention to cohort 2's intervention arm without adjustments. Harmful effects are defined as over 80% of the posterior credible interval in the harmful direction after the first cohort of N=240 participants. If this occurs, we will inform trial monitoring advisors immediately, and investigate potential causes with the local implementing partner.
2. If the posterior credible interval after the first cohort of N=240 participants demonstrate intervention effectiveness, defined as the 95% posterior credible interval in a protective direction which excludes the null, and the e-value corresponds to significant evidence (e-value >=20), then we will deem the intervention statistically effective. If the intervention is statistically effective **and** achieves greater cost-effectiveness on both outcomes compared with the evidence-based in-person intervention, we will continue the trial into the second cohort, which will involve delivering the current chatbot intervention to all participants in cohort 2 *(first to the intervention arm, then to the wait-list control arm, as the implementing partner does not have capacity to implement to all participants simultaneously).*
3. If the intervention is statistically effective **but does not** achieve greater cost-effectiveness on both outcomes compared with the evidence-based in-person intervention, we will collaborate with the local implementing partner to enhance the chatbot intervention delivery prior to cohort 2.
4. If the intervention is not statistically effective at the interim timepoint, but the posterior predictive probability of a protective effect after the end of the second cohort exceeds 95% **and** the intervention shows greater cost-effectiveness on both outcomes compared with the evidence-based in-person intervention, we will continue the trial into the second cohort, which will involve delivering the current chatbot intervention to all participants in cohort 2 *(first to the intervention arm, then to the wait-list control arm, as the implementing partner does not have capacity to implement to all participants simultaneously).*
5. If the intervention is not statistically effective at the interim timepoint but the posterior predictive probability of a protective effect after the end of the second cohort exceeds 95%, **but does not** show greater cost-effectiveness on both outcomes compared with the evidence-based in-person intervention, we will collaborate with the local implementing partner to enhance the chatbot intervention prior to the second cohort.

6. If the intervention is not statistically effective at the interim timepoint, **and** the posterior predictive probability of a protective effect after the end of the second cohort does not exceed 95%, we will collaborate with the local implementing partner to enhance the chatbot intervention delivery prior to the second cohort.
7. If the intervention is demonstrates promise for statistical significance (defined either as being statistically effective or having a posterior predictive probability of a protective effect with a second cohort exceeding 95%), **and** the intervention shows greater cost-effectiveness on both outcomes compared with the evidence-based in-person intervention, **but** programme completion rates are low, we will collaborate with our implementing partner to brainstorm how to strengthen programme completion prior to the second cohort.

5.6 Statistical Methods

Bayesian analyses will be conducted in multiple ways. We will use generalised linear models using the appropriate link function for each outcome, as done in our interim analysis in Ukraine (Tucker et al., 2026), and compare results with ordered logistic (OL) and partial credit (PC) models, using more advanced methods with the aim of achieving the best fit of the data. All results will be transparently shared, along with figures demonstrating which method provided the best fit of the data.

5.6 Missingness

If any outcome variable missingness is over 10%, we will use inverse probability weighting to account for missingness – via 'upweighting' participants with similar propensity scores of those who are missing.

5.7 Statistical Software

All data cleaning and analysis will be completed in R.

5.8 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.

5.6 Code

Intended code for estimating e-values is provided below. Full code from the interim analysis will be shared on Open Science Framework at the time of publication.

```
LAMBDA <- 1.5
```

```
M      <- 99999
M_SIG  <- 10000
set.seed(20260709)

contrasts <- function(m, n_t, k) {
  n <- length(m); s <- sum(m)
  vapply(seq_len(k), function(i) {
    ti <- sum(m[sample.int(n, n_t)])
    ti / n_t - (s - ti) / (n - n_t)
  }, numeric(1))
}

# benefit: +1 if higher is better, -1 if lower is better
e_outcome <- function(y, treat, benefit) {
  ok    <- !is.na(y) & !is.na(treat)
  y     <- y[ok]; treat <- as.logical(treat[ok])
  m     <- benefit * y / sd(y)
  n_t   <- sum(treat)
  sigma <- sd(contrasts(m, n_t, M_SIG))
  t_obs <- (mean(m[treat]) - mean(m[!treat])) / sigma
  t_pool <- contrasts(m, n_t, M) / sigma
  g_obs  <- exp(LAMBDA * t_obs)
  (M + 1) * g_obs / (g_obs + sum(exp(LAMBDA * t_pool)))
}
```

References

473 million Children Live in Conflict Zones – Peace Research Institute Oslo (PRIO). (2023). Retrieved 30 March 2026, from https://www.prio.org/comments/1152

Backhaus, S., Blackwell, A., & Gardner, F. (2025). The effectiveness of parenting interventions in reducing violence against children in humanitarian settings in low- and middle-income countries: A systematic review and meta-analysis. *Child Abuse & Neglect, Advancing Our Understanding of Violence against Children: Consideration of Global Threats of Conflict, Displacement, and Climate Change.*, *162*, 106850. https://doi.org/10.1016/j.chiabu.2024.106850

Cann, A., Calhoun, L. G., Tedeschi, R. G., Taku, K., Vishnevsky, T., Triplett, K. N., & Danhauer, S. C. (2010). A short form of the Posttraumatic Growth Inventory. *Anxiety, Stress, and Coping*, *23*(2), 127–137. https://doi.org/10.1080/10615800903094273

Carroll, H. A., Hook, K., Perez, O. F. R., Denckla, C., Vince, C. C., Ghebrehiwet, S., Ando, K., Touma, M., Borba, C. P. C., Fricchione, G. L., & Henderson, D. C. (2020). Establishing reliability and validity for mental health screening instruments in resource-constrained settings: Systematic review of the PHQ-9 and key recommendations. *Psychiatry Research*, *291*, 113236. https://doi.org/10.1016/j.psychres.2020.113236

Cianchetti, C., Pittau, A., Carta, V., Campus, G., Littarru, R., Ledda, M. G., Zuddas, A., & Fancello, G. S. (2013). Child and Adolescent Behavior Inventory (CABI): A New Instrument for Epidemiological Studies and Pre-Clinical Evaluation. *Clinical Practice and Epidemiology in Mental Health: CP & EMH*, *9*, 51–61. https://doi.org/10.2174/1745017901309010051

Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. (1998). *Psychological Medicine*, *28*(3), 551–558. https://doi.org/10.1017/s0033291798006667

Grantham, K. L., Forbes, A. B., Hooper, R., & Kasza, J. (2024). The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Statistical Methods in Medical Research*, *33*(1), 24–41. https://doi.org/10.1177/09622802231202364

Meinck, F., Boyes, M. E., Cluver, L., Ward, C. L., Schmidt, P., DeStone, S., & Dunne, M. P. (2018). Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. *Child Abuse & Neglect*, *82*, 45–58. https://doi.org/10.1016/j.chiabu.2018.05.022

Ramdas, A., Grünwald, P., Vovk, V., & Shafer, G. (2023). Game-theoretic statistics and safe anytime-valid inference. Statistical Science, 38(4), 576–601.

Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, *25*(3), 317–329. https://doi.org/10.1207/s15374424jccp2503_8

Tucker, S., Baldonado, N., Ruina, O. H., Melendez-Torres, G. J., Flaxman, S., Bryn, L., Louchnikov, K., Baldonado, J., Kharchenko, S., Machabelii, T., Goldman, P., Sherr, L., Ratmann, O., Lachman, J., Cluver, L., & Hillis, S. (2026). *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 No. 5228739). Social Science Research Network. https://doi.org/10.2139/ssrn.5228739

Tucker, S., Hawatmeh, N., Abu Lail, M., Lachman, J., & Hillis, S. (2025, January 26). Hope Groups: A Randomised Roll-Out Trial Of Psychosocial And Parenting Support Groups For Palestinian Caregivers Affected By War In the Middle East. https://doi.org/10.17605/OSF.IO/EWAQ4