Statistical Analysis Plan for SAATHI

SAATHI: School-based group interpersonal therapy for adolescents with depression in Nepal: a pilot realist cluster-randomised controlled trial

Trial Identifiers:

ISRCTN:	ISRCTN14652885		
Clincaltrials.gov	NCT06017700		
Protocol Version Number:	2.3	Date:	05/07/24
SAP Name	SAATHI_SAP_v0.1.docx	Date	04/04/2024

Names, affiliations, and roles of SAP contributors:

Statistician: Dr John Hodsoll, Department of Biostatistics and Health Informatics, IoPPN, King's College London

Chief Investigator: Kelly Rose-Clarke, UCL Institute for Global Health, University College London

SAP Version History

Date	SAP version	Protocol version and date	Reason for change/amendment made
	v1.0	v1.0	Initial version

Table of Contents

1. Trial Description

1.1 Brief description

We aim to assess the feasibility and acceptability of delivering group IPT in secondary schools in Chitwan Nepal, as well as the feasibility and acceptability of trial procedures. Further details of the background are available in the trial protocol, current version 2.3.

1.2 Aims and objectives

The objectives for the trial are as follows:

1. Is it feasible to deliver IPT in secondary schools in Chitwan?
2. Is the intervention acceptable to participants?
3. Is the intervention acceptable to caregivers and teachers?
4. Is it feasible to train and supervise local lay people to deliver IPT?
5. Are trial procedures (randomisation, blinding, data collection, safety standard operating procedures, control conditions) feasible to implement and acceptable to participants and schools?
6. How reliable are the measures for trial outcomes?
7. What are the recruitment, retention and response rates for trial participants?
8. How does context affect implementation of IPT?
9. Do data from the process evaluation support or refute theorized Context-Mechanism-Outcome-Configurations (COMCs)?
10. What are the costs per participant of implementing IPT in schools in Chitwan?

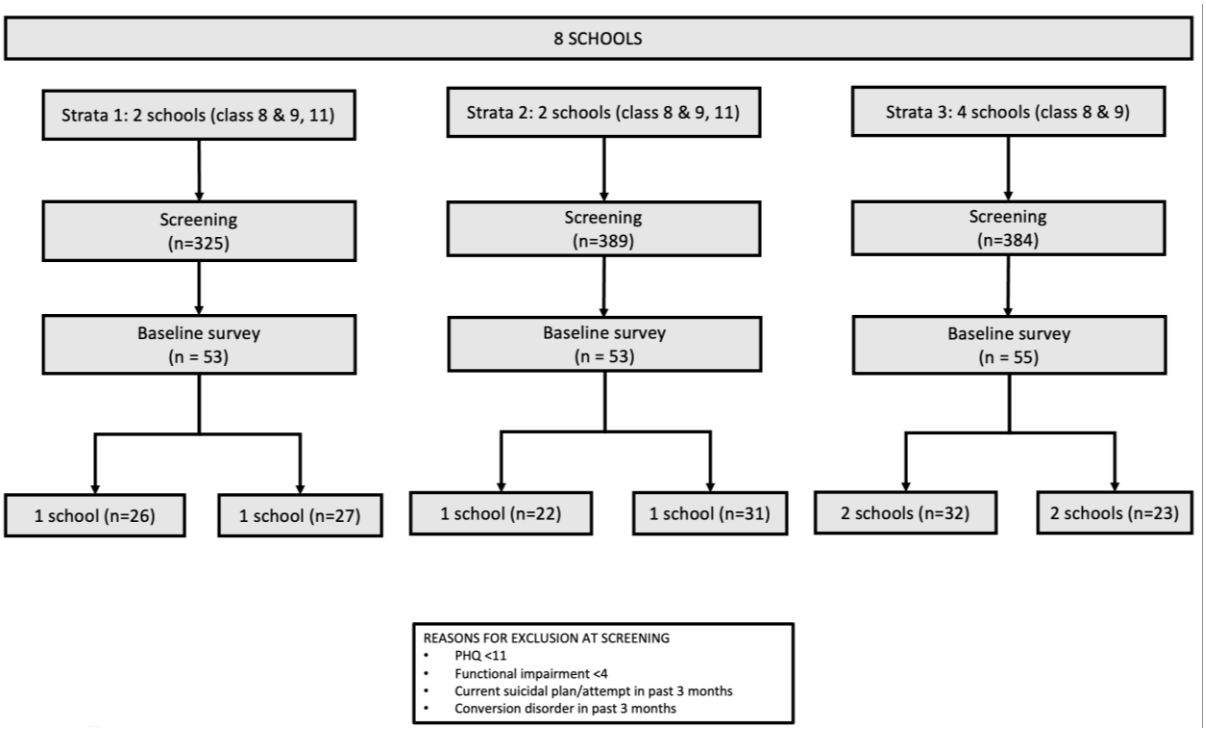
1.3 Study design.

We are doing a parallel group, two-arm superiority cluster-randomised controlled trial with 1:1 allocation ratio. The trial will be conducted in eight schools (four intervention and four control). We will assess participants in intervention and control arms at baseline, after the second group session (Midline 1), after the sixth group session (Midline 2), at endline (within two weeks of the final group session) and at follow-up (12 weeks after the final group session). The primary analysis in the phase III trial will be a cross-sectional comparison of mean depressive symptom scores at follow-up across trial arms, adjusted for baseline scores and potential confounders, and accounting for clustering.

1.3.1 Study setting

The study setting is Chitwan (चितवन), a mainly rural district in the lowland region of Nepal on the border with India. Within Chitwan, the pilot trial will run in schools in two rural municipalities, Rapti and Khairahani.

Figure 1. Trial design and flowchart



1.3.2 Trial Arms

1.3.2.1 Intervention

This is based on the WHO group IPT manual and incorporates modifications to the delivery model and content to enhance acceptability and effectiveness (1). Modifications include incorporating non-stigmatising Nepali mental health terminology; framing IPT as a life-skills training rather than mental health treatment to mitigate potential stigma towards participants; and using singing, dancing and storytelling to build relationships between group members and improve engagement.

The intervention involves two pre-group sessions and 10 group sessions. In the first pre-group session, the facilitator will meet the participant at school to identify the most relevant IPT problem area, help the participant link their depressive symptoms to the problem area, and gather information about the participant's key relationships and history of depression. In the second pre-group session, the facilitator will meet the participant and their caregiver together, ideally at home, to mobilise support and build rapport with the participant's family.

IPT groups are gender specific and comprise 6-8 participants per group. Facilitators will work in pairs with one facilitator designated as lead and the other as assistant. Group sessions will take place in a quiet, private space in the school (such as an empty classroom or the library). In the initial group session, the facilitator will focus on encouraging participants to review and share their problems, and instilling hope for recovery. In the middle sessions (2–9), participants will practice interpersonal skills,

and offer and receive support from group members to resolve their problems. In the last session, participants will review and celebrate progress, and make plans to tackle future problems.

In each pre-group and group session, participants will review their depressive symptoms with the facilitator using a seven-item symptom checklist developed for the study. This review process helps participants to link changes in symptoms to events in their daily lives and enables facilitators to identify deterioration and suicidality. We will implement a standard operating procedure to manage adolescents reporting suicidal thoughts, including risk assessment, consultation with an IPT supervisor, communication with parents and, where appropriate, one-to-one intervention for the adolescent with a psychosocial counsellor in parallel with the group sessions.

1.3.2.2 Control

Participants attending schools in the control arm will receive enhanced usual care. In intervention and control arms, we will train health workers in health posts and primary care centres using the WHO mental health GAP Action training package (mhGAP). Participants in the control clusters will receive a handout with information about the location of these trained health workers and how they can access treatment. Similarly to the intervention arm, we will implement a standard operating procedure to manage adolescents reporting suicidal thoughts, including risk assessment, communication with parents and, where appropriate, one-to-one intervention for the adolescent with a psychosocial counsellor in parallel with the group sessions. Participant in the control arm will not meet with an IPT facilitator.

1.4 Allocation method and blinding of groups

1.4.1 Unit of randomization

The trial is cluster randomised with the unit of randomization being school.

1.4.2 Randomization

Randomisation will be at the school level and stratified according to whether recruitment at the school includes adolescents from class 11 or not. In total there will be three strata: Strata 1 will comprise two of the four schools from which we are recruiting adolescents in class 8, 9 and 11. Strata 2 will comprise the two remaining schools in class 8, 9 and 11. Strata 3 will comprise schools where we are recruiting in class 8 and 9 only. We will complete screening and the baseline survey in Strata 1 (which will take approximately two weeks) then randomise these schools with a 1:1 allocation ratio to the intervention and control arms. The intervention will then start in these schools. After Strata 1 schools have been allocated, screening and baseline will start in Strata 2 after which these schools will be randomly allocated to the intervention and control arms. The process will then be repeated in Strata 3. In Strata 1 and 2 there are two schools with four groups per school (eight adolescents per group): one male and one female group for Class 8 and 9 combined, one male and one female group for Class 11. In Strata 3 there are four schools with two groups per school: one male and one female for Class 8 and 9 combined. Advantages for stratifying in this way include: (i) the duration between identifying that an adolescent is depressed and the adolescent starting the intervention is minimised; (ii) a staggered start to the surveys - meaning fewer research assistants are needed over the course of the trial; (iii)

randomisation can be conducted after baseline with minimal delay. The randomisation will be done by a blinded, independent statistician based in the Institute for Global Health, University College London.

1.4.3 Blinding

Research and clinical staff will work independently from separate offices in Chitwan. Research supervisors and the research assistants conducting baseline, midline, endline and follow-up surveys will not be given information about the allocation of schools. Due to the participatory nature of the intervention, adolescents, school staff, IPT facilitators and research assistants conducting the process evaluation will not be blind to allocation. The statistician conducting the final analysis of the pilot trial data will be blinded. Psycho-social counsellors available at participating schools will also be blinded.

1.4.4 Recruitment

1.4.4.1 Schools

Participants will be recruited from eight government secondary schools in Rapti and Khairahani Municipalities. These two municipalities were selected based on the consent received from the Mayor's Office. There are 14 government secondary schools in these municipalities. One school was excluded because it was already receiving services from TPO Nepal. From the remaining 13 we selected the eight that were easiest to access from the project office. In four schools we will recruit adolescents from class 8, 9 and 11. In the four remaining schools we will recruit adolescents in classes 8 and 9 only. Implementing groups with older and younger adolescents will enable us to explore the feasibility and acceptability of the intervention across these two age groups. Moreover, some of the schools we have selected are for class 1-10 only.

1.4.4.2 Participants

From all schools, we will obtain a register of adolescents in class 8 and 9. In Strata 1 and 2 schools we will also obtain a register of adolescents in class 11. In each school we will randomly order the adolescents on these registers. We will screen adolescents for eligibility according to this random order until we have sufficient participants. Adolescents who screen positive and meet the inclusion criteria will be given an information sheet about the trial and a consent form to be signed by themselves and their parents. Adolescents who return signed consent forms will be invited to participate in the trial.

1.5 Duration of the treatment period

The treatment period begins 4 to 5 weeks post-randomisation and lasts for 15 / 16 weeks.

1.6 Frequency and duration of follow-up

There are 5 measurement timepoints in total and 4 following randomisation:

1. The first intermediate timepoint after the first group session, at week 5.
2. The second intermediate timepoint will occur at week 12 after group session 6.
3. The endline timepoint is 20 – 21 weeks

4. Follow – up is at 45 – 46 weeks

1.7 Data Collection

Research assistants will mainly conduct screening, baseline, endline and follow-up interviews at school in a private place. If participants are not attending school regularly, research assistants will ask to conduct an interview in the participant's home. The interviews will last around one hour. In interviews, the research assistants will administer surveys to the participants primarily face-to-face in their school, outside of class hours. We will collect data on demographic characteristics (age, gender, religion, caste/ethnicity, level of education), socio-economic background (income sufficiency, main source of income), target primary (depression), secondary (anxiety, PTSD, functional impairment, school attendance and achievement) and intermediate outcomes (hope, self-esteem, emotion regulation, interpersonal conflict, interpersonal skill use, and social support for both arms, and group cohesion for intervention arm only) and predictors (school climate, gender norms, and adversity). We will use the Kobo Toolbox data collection platform. Research assistants will enter data on mobile phones or tablets. We will use automated skip patterns and consistency logic to reduce errors and missing data. Data on school attendance will be collected at baseline and endline from school registers. We will use school and telephone contact to try to follow up all participants. In the pilot trial, we will consider phone-based interviews for participants who move out of the area whom we cannot meet in person and describe any differences that may arise by mode of interview.

To pilot surveys with adolescents' mothers at baseline and endline, research assistants will conduct interviews at the mother's home or invite them to attend for an interview at their son/daughter's school. The interview will include questions on demographics (age, gender, religion, caste/ethnicity, level of education), socio-economic status, parenting skills, depressive symptoms and disruptive behaviour of their son/daughter.

During the surveys, the project coordinator will regularly download data from the server to check the number of interviews completed and identify any errors or missing data. We will pseudonymise the final dataset by removing personally identifiable information and store it on TPO Nepal's secure central server and KCL Sharepoint.

1.8 Eligibility criteria

1.8.1 Cluster selection

Participants will be recruited from government secondary schools in Rapti and Khairahani districts in Chitwan. Schools with more than 60 boys and 60 girls in class 8 and 9 combined, or more than 60 boys and 60 girls in class 11 alone were selected. This is because:

(i) groups are gender specific;
(ii) we estimate a prevalence of depression of around 10% and we need at least six adolescents per group;
(iii) findings from the feasibility study suggest adolescents in class 8 and 9 can be in the same group but that older adolescents in class 11 should be in a separate group.

(iv) Schools must also be easily accessible by road.

1.8.2 Participant Inclusion

Eligible participants will be:

- adolescent boys and girls aged 13-19;
- attending a participating school;
- enrolled in class 8, 9 or 11;
- with depression (i.e. scoring 11 or more on the PHQ-A); and
- with functionally impairment (i.e. scoring 4 or more on the functional impairment tool).

1.8.3 Participant Exclusion

Adolescents will be excluded if they are:

- in class 10 or 12 because they are busy preparing for School Education and Plus 2 exams;
- in class 7 and therefore may be too young to benefit from IPT;
- have suicidality (current plan or recent attempt in the past three months) because these adolescents require more acute, intensive treatment; or
- have experienced conversion disorder ("chhopne") in the past three months because group-based treatments may not be appropriate.

1.9 Outcomes

1.9.1 Indicators for feasibility outcomes

Table 1: Indicators and data sources for feasibility research questions. Grey text denotes analysis not under the remit of the trial statistician.

Question	Indicators and data source
1. Is it feasible to deliver IPT in secondary schools in Chitwan?	● Proportion of planned IPT sessions delivered. ● Focus group discussion and interview transcripts with IPT facilitators, adolescents, teachers and caregivers.
2. Is the intervention acceptable to participants?	● Proportion of adolescents approached who consent to participate in the trial. ● Proportion of IPT sessions attended by intervention participants. ● Proportion of all schools invited who agree to participate in the trial. ● Intervention participant treatment satisfaction surveys, percentage of participants rating treatment as helpful ● Focus group discussion and interview transcripts with intervention participants
3. Is the intervention acceptable to	● Focus group discussion and interview transcripts with caregivers and teachers

caregivers and teachers?	
4. Is it feasible to train and supervise local lay people to deliver IPT?	• Therapeutic competency assessed with ENACT and GroupACT tools. • Intervention fidelity assessed with the IPT checklist. • Proportion of facilitators trained who pass a paper-based IPT knowledge test. • Proportion of IPT sessions observed by supervisors. • Focus group discussion and interview transcripts with intervention participants and facilitators
5. Are trial procedures (randomisation, blinding, data collection, safety standard operating procedures, control conditions) feasible to implement and acceptable to participants and schools?	• Proportion of schools that consent to participate and are retained throughout the trial. • Proportion of eligible adolescents and schools that consent to participate in the trial • Baseline, midline, endline and follow-up survey response rates • Focus group discussion and interview transcripts with IPT facilitators, adolescents, teachers and caregivers. • Caregiver survey response rates • Rates of missing items on trial outcomes • Number and type of adverse event and response.
6. How reliable are the measures for trial outcomes?	• Research assistant inter-rater reliability for primary and secondary outcomes. • Internal consistency of each outcome measure
7. What are the recruitment, retention and response rates for trial participants?	• Proportion of eligible adolescents and schools that consent to participate in the trial. • Proportion of trial participants and schools retained at endline. • Trial participant baseline, midline, endline and follow-up survey response rates
8. How does context affect implementation of IPT?	• Recruitment, retention and response rates by gender, school, age • Focus group discussion and interview transcripts with IPT facilitators
9. Do data from the process evaluation support or refute CMOCs?	See Protocol Table 5.
10. What are the costs per participant of implementing IPT in schools in Chitwan?	• Cost data • Time sheets • Interviews with SAATHI-2 team members

1.9.2 Feasibility outcomes for progression criteria

The recruitment, implementation, fidelity and retention of the pilot trial will be evaluated with respect to the criteria green / amber / red; green means progress, amber flagged for further evaluation and red means do not progress. Any one indicator flagged as red would be sufficient to stop the trial. The feasibility outcomes are listed in the table below.

Research question	Criterion	Indicator	Green	Amber	Red
Is it feasible to deliver IPT in secondary schools in Chitwan?	Intervention implementation	Percentage of planned IPT sessions delivered	> 70%	40-70%	< 40%
Is the intervention acceptable to participants?	Participant treatment satisfaction	Percentage of participants rating IPT as 'quite helpful' or 'very helpful';	> 67%	30-66%	< 30%
	Treatment adherence	Percentage of participants who attend more than 70% of IPT group sessions	> 50%	20-50%	< 20%
Is it feasible to train local lay people and school nurses to deliver IPT?	Fidelity to IPT	Percentage of session components rated superior or satisfactory, averaged across observed sessions	> 60%	30-59%	< 30%
	Serious adverse events	Difference in serious adverse events in the intervention arm compared to the control arm	<4	4	5
Are trial procedures (randomisation and data collection) are acceptable to participants and schools?	Eligible adolescents agree to participate	Percentage of adolescents with informed consent at baseline**	> 60%	30-60%	< 30%
	Eligible schools agree to participate	Percentage of schools approached that agree to participate	> 60%	30-60%	< 30%
	Missing data	Percentage of missing items on primary and secondary outcome	< 15%	15-50%	> 50%
	Participant retention	Percentage of participants completing the endline survey	> 70%	30-70%	< 30%

***NOTE: Current study protocol does not have a separate consent procedure after randomization. Adolescents and parents give consent for both screening and the full trial. TBD to be determined.*

1.9.3 Primary Outcome

- The potential primary outcome for the phase 3 trial is the adolescent version of the Patient Health Questionnaire (2) translated for Nepal (PHQ – A) (3).

1.9.4 Secondary Outcomes

- Functional impairment; Locally developed tool, an 11-item tool which is adapted for Chitwan and for boys and girls (4)
- Anxiety; Generalised Anxiety Disorder Assessment (GAD-7) (3)
- Post-traumatic stress disorder; PCL-C (8-item abbreviated version of the PTSD Checklist) based on the DSM-5 symptoms of PTSD. (5)
- School attendance; Number of days absent in past 12 days (excluding school closures) which will be collected from the school register
- Educational performance; scores on the end of year school examination
- Quality of Life; EuroQol-5 Dimension 5 levels (EQ-5D) is a self-report measure of health related quality of life and can be used to determine quality adjusted life years (6).

1.9.5 Intermediate outcomes

- Hope: Children's Hope Scale is a self-report measure of hope in individuals age 8 to 16, evaluating their perceptions that their goals can be met. It is a 6 item likert-scale with 6 levels on each item ranging from "none of the time" to "all of the time". There are 2 domains; pathways to hope and agency. (7)
- Emotion regulation: Difficulties in Emotion Regulation Scale (DERS-SF) 18 items (8)
- Self-efficacy: General Self-efficacy Scale 10 item scale
- Social support: Multidimensional Scale of Perceived Social Support
- Group cohesiveness: Group Cohesiveness scale
- IPT skills: Interpersonal Psychology Skills Scale
- Conflict reduction: Social Adjustment Scale Self Report

1.9.6 Other outcomes

- School climate; Beyond Blues scale (10 items selected from 28 original items) (9)
- Gender norms; Johns Hopkins Global Early Adolescent Study - 9 items
- Adversity; Johns Hopkins Global Early Adolescent Study 13 items
- Mother's depression; Patient Health Questionnaire, 9 items (10)
- Parenting; Alabama Parenting Questionnaire 19 items
- Disruptive behaviour; Disruptive Behavior International Scale – Nepal version (DBIS-N) 10 items

1.9.7 Process indicators

In intervention clusters, we will collect data on intervention fidelity, context, mechanisms and dose using; competency checklists (ENACT, GroupACT, IPT checklist), unstructured observation of group

sessions, notes on facilitator supervisions with clinical supervisors and attendance registers from group sessions.

For each IPT group a supervisor will observe the initial session, two sessions from the middle phase and the termination session (four sessions per group). A quantitative score for intervention fidelity will be generated using the IPT checklist, a checklist of key session components that should be carried out by the facilitator (e.g. discusses confidentiality, outlines group rules, works to establish rapport, and skills related to the IPT problem areas). Whilst observing IPT sessions, the supervisor rates each component as superior, satisfactory, needs improvement, or failed to attempt. We will calculate fidelity as the percentage of session components rated superior or satisfactory, averaged across observed sessions.

At endline we will conduct focus group discussions (FGDs) and interviews to explore possible mechanisms of IPT and contextual moderators of IPT's effects with facilitators, teachers, caregivers and adolescents. This is further described in the trial protocol and will not be analysed by the trial statistician.

1.10 Sample size

The pilot trial is not powered to detect an effect of the intervention on depression. In Strata 1 and 2 schools we will recruit eight girls and eight boys from class 8 and 9, and eight girls and eight boys from class 11. In Strata 3 schools we will recruit eight girls and eight boys from class 8 and 9 only. This gives a total sample size of (n=192). In each of the Strata 1 an 2 schools allocated to the intervention we will pilot four IPT groups of eight adolescents (one group for boys in class 8 or 9, one for boys in class 11, one for girls in class 8 or 9 and one for girls in class 11). In Strata 3 schools we will pilot two IPT groups of eight adolescents (one group for boys in class 8 or 9 and one for girls in class 8 or 9). The decision to include eight clusters was informed by the available budget and resources. Although we will estimate the intra cluster correlation coefficient for depression and the recruitment rate using the baseline data, these will be biased downwards due to the small number of clusters [23]. To calculate the sample size for the full trial, we will triangulate estimates from the pilot trial with estimates from other school-based trials in Nepal and similar settings.

2. Descriptive Statistics

2.1 Flow of participants

The CONSORT flow chart will be compiled in accordance with the criteria of the Consolidated Standards of Reporting Trials (CONSORT) 2010 extension for cluster-randomised trials (6). This will include information at both the cluster and individual level, including the number of eligible clusters and of those, the number agreeing to participate in the trial. Similarly at the participant level, we will report the number of students eligible and the number agreeing to enter or refusing to enter the trial. Then by trial arm, number of participants followed through the trial, the number withdrawing and the number lost to follow-up.

2.2 Baseline comparability of randomised groups

Baseline demographic and clinical variables will be summarised for each of the two arms of the trial (and across trial arms) at the cluster and individual participant level. Means and standard deviations or medians, minimum, maximum and interquartile range for continuous measures and frequencies and proportions for binary, ordinal or multinomial categorical measures. Although randomization ensures any differences between clusters are by chance, this means that there may be differences at the individual level. No tests will be performed to compare the groups but effect sizes for between-group differences will be estimated (according to data type) and reported.

2.3 Intercurrent or intermediate events

2.3.1 List of potential events

We summarize a list of potential intercurrent or intermediate events. These are events happening between randomisation and collection of final outcome data. These are potentially important as they might affect interpretation of the causal effect of treatment through adherence to the intervention or loss to follow up. Where possible we will describe the frequency of these events by treatment arm noting that some overlap with adverse events. Potential intercurrent events are:

1. Non-adherence to treatment (treatment discontinuation, missed sessions or not starting treatment).
2. Discontinuation of participation due to an SAE
3. School-wide event such as closure due to extreme weather / heat, examinations
4. Disruption of the group by individual conflict
5. Significant life events such as family issues.
6. Death: extremely unlikely to happen and be related to the trial

2.3.2 Handling of intercurrent events:

We will collect and include all data from all cases regardless of whether intercurrent events related or unrelated to the treatment or disease) happened or not. Intercurrent events will be reported in terms of frequencies and proportions as needed.

2.4 Loss to follow-up and other missing data

Rates of withdrawal will be summarized for adolescents using frequencies and proportions. The reasons for withdrawal from treatment and from further data collection will be summarised.

The proportion of participants missing each outcome variable at follow-up will be summarised in each arm. If available, the reasons for missing data at baseline and at follow-up will be summarised within each trial arm. It will be ensured that the patterns of missing data and reasons for missing data are consistent with the CONSORT flow diagram. We will report patterns of missing data, i.e. those variables that are frequently missing concurrently and the frequency of each missing data pattern.

To evaluate whether there are predictors of missingness at baseline, summaries of baseline variables will be presented separately and compared for those completing the trial and those being lost to follow-up. Binary logistic regression will be used to assess possible predictors of missingness potentially for inclusion in imputation models.

2.4.1 Missing items in scales

Missing data at the item level will be handled by questionnaire-specific recommendations where available. Otherwise, data for an individual will be pro-rated if 20% or fewer items are missing. For example, in a scale with 10 items, individuals with only 1 or 2 items missing will have those missing scores pro-rated. The average value for the 8 or 9 complete items will be calculated for that individual and used to replace the missing values. The scale score will be calculated based on the complete values and these replacements. Simulation studies have shown that pro-rating is robust method when data are missing on less than 20% of items (11).

2.5 Adherence to allocated treatment

Adolescents that are compliant with the treatment (i.e., those who received more than 70% of the group sessions) versus non-compliant will be described including rates of treatment withdrawal. As for loss to follow-up, reasons for non-adherence will be summarized per treatment arm and logistic regression will evaluate baseline associations of adherence.

2.6 Feasibility outcomes

Trial feasibility will be assessed according to the outcomes and progression criteria listed in sections 1.10.1 and 1.10.2. Feasibility parameter estimates will be accompanied by 95% confidence intervals

(CI) to provide a measure of estimator precision. Summaries will be produced using proportions, means and standard deviations or median and interquartile range as appropriate.

2.7 Harms

Information on potential harms will be collected through the qualitative interviews and FGDs in the process evaluation. Any serious adverse events (e.g. a suicidal attempt or life-threatening injury) will immediately to the Trial Steering Committee (TSC) who will decide what action should be taken. We will report other adverse events to the TSC at follow-up meetings.

2.8 Descriptive statistics for outcome measures

We will summarise measures of the primary and secondary outcomes listed under section 1.9 at baseline and endline at the individual level (i.e. giving equal weight to either cluster or individual respectively). Depending on the distribution of the measure, these will be either means and standard deviations or medians, minimum, maximum and interquartile range or proportions and 95% confidence intervals. Distributions will be visualised with boxplots violin plots and histograms for continuous or scale variables and barplots for discrete or categorical variables.

3. Data analysis – Clinical outcomes

3.1 Primary analysis of trial arm differences

The primary objective of the analyses is to estimate the mean group difference in mean depression symptoms (measured by the PHQ – A) for adolescents aged 13-19 at endline for the IPT group versus TAU. This will be estimated at the participant level as we are interested in the impact of IPT on individuals. Further the group difference will be estimated irrespective of whether participants discontinued IPT as we wish to know the effect of IPT as part of potentially routine care in the real world. Mean group differences and 95% confidence intervals adjusted for clustering will be reported as the estimate of treatment efficacy.

We define the estimand as follows:

- **Population:** 13- to 19-year-old adolescents from a randomised cluster (school) who meet the trial inclusion / exclusion criteria and are identified as depressed by the PHQ-A.
- **Outcome:** Depression scores following post-randomisation
- **Intervention:** IPT compared to TAU irrespective of treatment being discontinued (ITT or treatment policy strategy).
- **Population level-summary measure:** Mean PHQ – A at endline for each treatment group.
- **Comparison:** Mean group difference between treatment groups at the participant-level
- **Intercurrent events:** Interruption or discontinuation of treatment is ignored in the treatment policy strategy. The pattern of withdrawals in each treatment group will be assessed and any necessary remedial action will be approved by the TSC.

3.1.1 Effect size for primary outcome

For the primary outcome we will take a linear mixed model ANCOVA approach, adjusting for the statistical dependency resulting from within school clustering, repeated measures per student and for student baseline depression scores. This model is estimated using restricted maximum likelihood (REML) and will allow inclusion of all participants with missing data but at least 1 measure outcome observation under the assumption of missing at random (MAR, see section 3.x). We will include both endline and follow-up in the model which will have the following fixed effects: time (endline vs follow-up), treatment group, a time by treatment group interaction to allow the estimation of the effect of treatment group at both endline and follow-up and finally municipality as the stratification factor. In addition to the fixed effects, random effects of cluster and individuals (due to repeated measures at endline and follow-up) will be entered into the model to allow for correlations between the individual participants of each cluster. A random effect of time at the cluster level to allow within time correlations to be stronger within than across time. Finally as well as the baseline score for individuals, the baseline score per cluster can be included as a fixed effect (13)

As this is a pilot trial, the focus of the analysis will be effect sizes and 95% confidence intervals both standardized and unstandardized. Standardized effect sizes will be calculated by dividing the treatment group difference estimate by the baseline standard deviation (14). As the standard deviation is an estimate rather than a known population parameter, bootstrapping (wild bootstrap,

see section 3.2.x below) will be used to generate 95% confidence intervals. The model will be fitted in R 4.3 using the lme4 package. The code will be as follows:

```
lmer(phq ~ arm * timepoint + phq0 + municipality + (timepoint || school) + (1 | id),
        data = saathi_df)
```

phq – PHQ-A scores at endline and follow-up
arm – treatment arm
timepoint – endline versus follow-up
phq0 – PHQ-A scores at baseline
municipality – municipality of school
school – school identifier
id – participant identifier

3.1.2 Checks for model assumptions

Following model fitting, the assumptions of the linear mixed effects models will be examined. Firstly, it assumes the residuals are approximately normally distributed. Residuals in each cluster and at each time-point will be plotted using a Q-Q plot to check for normality. Bootstrapped 95% confidence intervals will be calculated by default which will allow some flexibility if model assumptions are violated.

3.1.3 Secondary and intermediate outcomes

The secondary and intermediate outcomes are all continuous with identical estimands apart from the outcome. Therefore, the same analysis models (and assumption checks) will be used to assess treatment effects on these outcomes as for the primary analysis.

3.1.4 Per protocol analysis

To assess the effects of receiving the treatment according to the definition in the trial protocol, we will estimate effect sizes according to whether participants received an acceptable “dose” of the intervention or not. A binary exposure variable to identify those participants who attended more or equal to 7 group sessions (1 = yes and 0 = no). The effect of adherence will be estimated by including the interaction between the exposure variable and treatment group, noting that this approach does have limitations such as assuming there is no unmeasured confounding. A CACE analysis will be considered if non-compliance is above 20%.

3.2 Statistical considerations

3.2.1 Outcome distribution

The scale for the primary outcome PHQ-A generally shows a skewed distribution in the general population, but given the inclusion criteria may reasonably approximate a normal distribution (15) and allow linear models to be used for analysis. If the model error distribution does not accord to

model assumptions due to skew, as stated in section 3.1.2, bootstrapping is the default approach for confidence intervals. An alternative for right-skewed distribution typical of depression outcomes is the negative binomial model, which may be considered in a sensitivity analysis.

3.2.2 Clustering

As the unit of randomisation in this study was schools, the statistical model for the analysis of the data will need to take into this clustering, as individuals within clusters are more likely to be similar then individuals across clusters. Analysis options in this scenario include generalized linear mixed models (GLMM) including subject specific random effects by cluster or generalized estimating equations (GEE), where an overall or working correlation matrix describes the statistical dependence between individuals in a cluster. Here GLMM will be used as this approach allows a more flexible approach, for example allowing the effect of time to vary across clusters. As the focus of treatment is the individual (albeit in group therapy) we are interested in treatment effects at the participant level (section 3.1). Note that in GLMMs the marginal and specific cluster / participant effects can differ depending on the distribution (and link function) used to model the outcome. However, note that the marginal and participant specific effects will be the same for the primary outcome given the use of linear mixed models for statistical inference.

Another issue is whether cluster size is informative in terms of the treatment effect, i.e. the intervention could be more or less effective depending on the size of the cluster (16). With a participant-level analysis, each participant should receive equal weight in the analysis. However, both mixed model and GEE analyses weight participant information by inverse variance of the cluster. As In this case (16) suggest using a GEE with independence working correlation matrix and cluster robust errors.

One challenge for the SAATHI study is the relatively small number of clusters at 8, which can lead to increase the risk of a type 1 error. (17). Whilst we will not present formal inference tests here, the same bias would mean the 95% confidence interval for the effect size was too small. The general solution for mixed models is to add a small sample correction by adjusting the degrees of freedom and for GEEs the solution is to correct the standard errors (SE). This is similarly the case with cluster robust standard errors being biased when the number of clusters is small. The subcluster wild bootstrap allows more correct coverage in this case (18).

3.2.3 Adjustment for covariates

There will be adjustment for baseline depression scores and the stratification factor municipality.

3.2.4 Missing outcome data

Linear mixed models will be estimated using ML. Such an approach provides valid inferences under the assumption that the missing data mechanism is ignorable (or missing at random, MAR). Here this means that explanatory variables included in the model can predict missingness as can earlier

observed values of the outcome. We will include baseline predictors of drop-out as explanatory variables in the linear mixed model to make the MAR assumption more realistic.

3.3 Exploratory analyses of CMOCs

Additional analyses related to the following CMOCs identified in the trial protocol are described in this section. The statistical approach to the analysis of CMOCs will be using descriptive statistics and including moderators and mediating pathways in statistical models. Whilst the association of context mechanism and outcomes is best identified as mediated moderation in which candidate context variables will moderate the relationship between treatment and mediators (the intermediate outcomes) identifying the putative mechanisms by which IPT works, the sample size here precludes fitting models of this complexity and indeed even simple moderation / mediation models unless large effect sizes are expected. Where statistical models are estimated they will be for overall moderation or mediation effects relevant to the CMOC and fit using a Bayesian approach which will allow the use of weakly informative priors to constrain estimates of the effect size to plausible values.

3.3.1 CMOC 1 – effect of SES

This is defined as follows:
Context; For students of higher socioeconomic status not experiencing structural vulnerability or intractable adversity and having more opportunities to implement new strategies.
Mechanism: IPT enables participants to participate in discussions through which they learn and implement strategies to develop hope, reduce conflict, build relationships and improve self-efficacy.
Outcome: This generates reductions in depression.

To assess this CMOC the following procedure will be carried out; to operationalize context a variable coding SES (high vs low) and exposure to adversity will be generated. Quantitative evaluation of process indicators and group differences for (1) attendance and (2) items 1 to 4 on group cohesion scale will be described (mean difference and 95% CI). The ability to implement IPT by SES will be estimated by group comparisons of measures of -self efficacy, hope, conflict and relationship initiation / improvement at midline points 1 and 2 in the treatment group only.

The difference in ES for the primary outcome (PHQ-A) will be estimated by adding an interaction term for SES by treatment group to the primary outcome model. Given the small sample size a Bayesian model with informative priors can be used to restrict the interaction term to a range of plausible values.

3.3.2 CMOC 2 – school setting

Context: For students in schools with school climates characterised by strong student-staff and student-student relationships and norms of mutual respect and social support and thereby having more opportunities to implement new strategies.
Mechanism: supportive for implementation or amenable for adolescents to implement IPT skills
Outcome: This generates reductions in depression.

For this CMOC context is determined by scores on the Beyond Blue scale (continuous). Effects on IPT processes can be evaluated by a positive spearman's rank correlation (and 95% CI) between BB scores and IPT indicators, self-efficacy, hope, conflict and relationship initiation / improvement in the intervention group. As for CMOC 1, differential effects of school atmosphere can be evaluated by adding an interaction term with school setting to the primary analysis model with a recommendation of a Bayesian approach.

3.3.3 CMOC 3 - Gender

Context: For boys who do not experience cultural norms or structural violence which impede opportunities to implement new strategies.
Mechanism: IPT enables participants to participate in discussions through which they learn strategies to develop hope, reduce conflict, build relationships and improve self-efficacy.
Outcome: This generates reductions in depression.

The context is determined by adolescent gender. Effects on IPT processes can be evaluated by a mean group (girls versus boys) differences IPT indicators, self-efficacy, hope, conflict and relationship initiation / improvement. As for CMOC 1, differential effects of gender can be evaluated by adding a gender by treatment group interaction to the primary outcome model.

3.3.4 CMOC 4 - Age

Context: For older (and more cognitively able?) students more cognitively able to consider others' perspectives, learn negotiation skills, develop solutions to interpersonal problems, understand links between events and mood, manage anger and name and express emotions
Mechanism: IPT enables participants to participate in discussions through which they learn strategies to develop hope, reduce conflict, build relationships and improve self-efficacy.
Outcome: This generates reductions in depression.

The variable defining the context is student age (group or continuous?). Better acquisition of IPT skills will be assessed by comparison of IPT knowledge test and IPT skills scores, emotion regulation (DERS) by age. Depression outcomes (PHQ-A) by age can be evaluated by including an interaction term for age and treatment group in the primary analysis model.

3.3.5 CMOC 5 – Early change

Context: For students participating in initial sessions and participating in discussions which validate participants' experiences and instil hope.
Mechanism: IPT enables participants to develop hope and motivation to engage in the intervention.
Outcome: This generates immediate reductions in depression.
The context here is early change which is operationalized by the first 2 therapy sessions. Early change can be assessed at the therapy session symptom checklist scores. The temporal relationship between hope and PHQ-A score can be described visually by plotting baseline, midline and endline data.

4. Data management

All data collection is being conducted in Nepal by TPO Nepal employees who have no employment affiliation with KCL. KCL's role is that there is a KCL co-PI (Dr Kelly Rose-Clarke) who will be receiving de-linked pseudo-anonymised datasets which will be analysed at KCL and stored on KCL servers. Dr Joanna Morrison, a co-I at UCL) will also be receiving de-linked pseudo-anonymised data which she will analyse.

TPO Nepal researchers will collect quantitative data from participants using mobile phones and/or tablets, owned by TPO Nepal, which will be password protected and encrypted where possible. They will unlink the identifiers from the dataset using unique participant IDs. The dataset containing identifiers will be stored on a password-protected file on TPO's secure central server.

TPO Nepal researchers will make audio recordings on voice recorders. Recordings will be uploaded to the TPO server (and deleted from the recorder) and deleted from the server once they have been transcribed or, where transcription is not required (e.g., transect walks, cognitive interviews), after detailed notes have been taken. Transcriptions and notes will be stored on the TPO Nepal server. Any paper data (e.g., hand-written notes) will be stored in a locked filing cabinet at the TPO Nepal office.

All personal data will be stored at TPO Nepal and not shared with KCL. Dr Rose-Clarke (PI) will only receive de-linked pseudoanonymised datasets, transferred using a secure file transfer and storage service.

5. Analysis software

STATA 18 and R 4.3 will be used for the main descriptive and inferential analyses

6. References

1. Rose-Clarke K, Pradhan I, Shrestha P, B.K. P, Magar J, Luitel NP, et al. Culturally and developmentally adapting group interpersonal therapy for adolescents with depression in rural Nepal. BMC Psychol. 2020 Aug 12;8(1):83.

2. Johnson JG, Harris ES, Spitzer RL, Williams JBW. The patient health questionnaire for adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolesc Health. 2002 Mar 1;30(3):196–204.

3. Luitel NP, Rimal D, Eleftheriou G, Rose-Clarke K, Nayaju S, Gautam K, et al. Translation, cultural adaptation and validation of Patient Health Questionnaire and generalized anxiety disorder among adolescents in Nepal. Child Adolesc Psychiatry Ment Health. 2024 Jun 19;18(1):74.

4. Bolton P, Tang AM. An alternative approach to cross-cultural function assessment. Soc Psychiatry Psychiatr Epidemiol Int J Res Soc Genet Epidemiol Ment Health Serv. 2002;37(11):537–43.

5. Price M, Szafranski DD, van Stolk-Cooke K, Gros DF. Investigation of abbreviated 4 and 8 item versions of the PTSD Checklist 5. Psychiatry Res. 2016 May 30;239:124–30.

6. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013 Sep 1;22(7):1717–27.

7. Snyder CR, Hoza B, Pelham WE, Rapoff M, Ware L, Danovsky M, et al. The Development and Validation of the Children's Hope Scale1. J Pediatr Psychol. 1997 Jun 1;22(3):399–421.

8. Gouveia P, Ramos C, Brito J, Almeida TC, Cardoso J. The Difficulties in Emotion Regulation Scale – Short Form (DERS-SF): psychometric properties and invariance between genders. Psicol Reflex E Crítica. 2022 May 6;35(1):11.

9. Sawyer MG, Pfeiffer S, Spence SH, Bond L, Graetz B, Kay D, et al. School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative. J Child Psychol Psychiatry. 2010;51(2):199–209.

10. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med. 2001 Sep 1;16(9):606–13.

11. Roth PL, Switzer FS, Switzer DM. Missing Data in Multiple Item Scales: A Monte Carlo Analysis of Missing Data Techniques. Organ Res Methods. 1999 Jul 1;2(3):211–32.

12. Phillips R, Cro S, Wheeler G, Bond S, Morris TP, Creanor S, et al. Visualising harms in publications of randomised controlled trials: consensus and recommendations. BMJ. 2022 May 16;377:e068983.

13. Hooper R, Forbes A, Hemming K, Takeda A, Beresford L. Analysis of cluster randomised trials with an assessment of outcome at baseline. BMJ. 2018 Mar 20;360:k1121.

14. Hopkins WG, Rowlands DS. Standardization and other approaches to meta-analyze differences in means. Stat Med [Internet]. [cited 2024 May 23];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.10114

15. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes. 2007 Nov 27;5(1):63.

16. Kahan BC, Li F, Copas AJ, Harhay MO. Estimands in cluster-randomized trials: choosing analyses that answer the right question. Int J Epidemiol. 2023 Feb 1;52(1):107–18.

17. Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: which analyses should be used? Int J Epidemiol. 2018 Feb 1;47(1):321–31.

18. MacKinnon JG, Webb MD. The wild bootstrap for few (treated) clusters. Econom J. 2018 Jun 1;21(2):114–35.