IMPlementation of facility and community interventions to reduce the treatment gap for dep**RESS**ion (IMPRESS)

TRIAL STATISTICAL ANALYSIS PLAN

Protocol Title	Effectiveness and Cost-effectiveness of a Community Intervention in Enhancing Access to Care and Improving Clinical Outcomes for Depression: A Protocol for a Cluster Randomised Controlled Trial in India
Protocol version	Version 0.10 May 2025
Trial registration number	ClincalTrials.gov (NCT05890222)
SAP version	Version 9 (June 2026)
Principal Investigator (PI)	Abhijit Nadkarni and Vikram Patel
Written by:	Abhijit Nadkarni
Reviewed by:	Helen Weiss (Co-I) and Vikram Patel
Signed off by:	Abhijit Nadkarni (PI), Vikram Patel (Co-PI), Helen Weiss (Co-I and Trial Statistician), Chunling Lu (Co-I), Daisy Singla (Co-I), Richard Velleman (Co-I)

Revision history

Version	Date
1.0	17 August 2025
2.0	17 October 2025
3.0	28 November 2025
4.0	6 May 2026
5.0	23 May 2026
6.0	27 May 2026
7.0	29 May 2026

8.0	16 June 2026
9.0	20 June 2026

Table of Contents

1 Introduction 5

1.1 Background and rationale 5

1.2 Objectives and hypotheses 6

2 Study Design 7

2.1 Trial design 7

2.2 Changes to the design and analyses 7

2.3 Sample size 11

2.4 Randomisation 13

2.5 Masking 14

2.6 Timing of outcome assessments and analysis 14

3 Trial population 14

3.1 Eligibility criteria 14

3.2 Participant recruitment 15

3.3 Participant recruitment at baseline 15

3.4 Participant follow-up 15

3.5 Analysis populations 16

4 Outcomes 16

4.1 Data sources 16

4.2 Primary outcome definitions 17

4.3 Secondary outcome definitions 18

5 Statistical analyses 21

5.1 Baseline characteristics and comparability by arm 21

5.2 Analysis of continuous outcomes 22

5.3 Analysis of binary outcomes 23

5.4 Analysis of outcomes with repeated measures per participant 24

5.5 Effect-modification and mediation 24

5.6 Sensitivity and additional analyses 26

5.7 Exploratory analyses 27

5.8 Economic evaluation 27

6 Statistical considerations 28

6.1 Missing data 28
6.2 Multiple testing adjustment 29
6.3 Intra-cluster Correlation Coefficient (ICC) and coefficient of variation 22
7 AEs and SAEs reporting 29
8 Software 31
9 References 32

1 Introduction

1.1 Background and rationale

Depression is the leading cause of disability worldwide, and accounts for more than a third of Disability Adjusted Life Years (DALYs) attributed to mental and substance use disorders[1]. The COVID pandemic substantially increased the prevalence of depression with an estimated additional 53·2 million cases of major depressive disorder globally (an increase of 27·6%) after the onset of the pandemic[2]. In 2017, 45·7 million individuals had depressive disorders in India, contributing the most (33·8%) to the DALYs related to mental disorders in the country[3].

Depression treatment coverage across the world is inadequate. It ranges from 33% in high-income countries to 8% in low- and lower middle-income countries (LMICs)[4]. Further, coverage of minimally adequate treatment (i.e. treatment levels deemed minimally sufficient for common mental health problems) is lower still, from 23% in high-income countries to 3% in LMICs[4].

In India, the District Mental Health Program (DMHP) was started in 1996 under the Indian National Mental Health Program (NMHP) to decentralise mental health services by integrating them in general health care delivery and at the community level. However, only a third of Indian states have more than 50% of the population covered by the DMHP, and the treatment gap for depression remains high at 85%[5].

Supply-side and demand-side barriers perpetuate this treatment gap. Supply-side barriers include the lack of skilled human resources. In low-resource settings, task-sharing with non (mental health) specialist workers has become one of the most widely implemented innovations to address this barrier[6]. The evidence on task-sharing of psychosocial treatments for depression is robust and indicates high levels of acceptability and effectiveness of brief psychosocial treatments, delivered over 6-12 weeks in 6-10 sessions by non-specialist workers through routine care and community platforms[7]. Demand-side barriers, notably the lack of recognition of the distress associated with depression as being a health problem, which combined with the stigma attached to mental health problems and help-seeking, leads to low levels of demand for and adherence to

depression care[8]. This results in both low *contact coverage* and low *effectiveness* for depression.

1.2 Objectives and hypotheses

Through the IMPRESS (IMPlementation of evidence based facility and community interventions to reduce the treatment gap for depRESSion) programme we aim to increase both the contact coverage and effectiveness of a scaled-up, evidence-based psychosocial treatment for depression (Healthy Activity Program-HAP)[9] and, in doing so, reduce the treatment gap for depression in a cost-effective manner. We will do this by examining if the addition of a community intervention (Community Model) delivered by community volunteers enhances the population-level impact of scaled-up HAP in routine primary care (Facility Model) by increasing demand for HAP and improving HAP adherence and effectiveness. We will conduct an economic evaluation, to estimate the cost-effectiveness of the Community Model in comparison to the Facility Model, and to use the estimates of the cost of the delivery of the interventions to model the financial sustainability of scaling it up across India.

(1) Our *primary hypotheses* are that the Community Model will be superior to the Facility Model in reducing the treatment gap for depression through:

(1a) increasing the demand for depression treatment in primary care (contact coverage outcome); and

(1b) reducing depressive symptoms at three months post-recruitment amongst those who screen positive for moderately severe and severe depression (effectiveness outcome).

(2) Our *secondary hypotheses* are that the Community Model will be superior to the Facility Model in

(2a) achieving a sustained reduction in symptoms of depression;

(2b) increasing short-term and sustained remission from depression;

(2c) achieving a greater response to HAP;

(2d) increasing perceived social support;

(2e) achieving higher treatment completion rates; and

(2f) increasing patient-reported activation levels.

In addition, we also hypothesize that the Community Model will increase community awareness about depression.

(3) Finally, we hypothesise that the superior outcomes of the Community Model will be mediated through:

(3a) increased community awareness about depression (for contact coverage);

(3b) greater perceived social support (for effectiveness);

(3c) higher treatment completion rates (for effectiveness); and

(3d) increased patient-reported activation levels (for effectiveness).

2 Study Design

2.1 Trial design

We conducted a Hybrid Type 2 Effectiveness Implementation Cluster Randomised Controlled Trial (CRT)[10], with equal allocation of clusters between the two arms. This design combined clinical effectiveness testing alongside implementation testing, thus aiding in better translation of research into practice, and improved and potentially effective implementation strategies. Our implementation evaluation component was based on the RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) framework[11]. This framework, used extensively in implementation research, enables the measurement of a logical flow of outcomes, better facilitation of translation of research findings, and focuses on the external validity of research findings[11].

2.2 Changes to the design and analyses (approved by IRB)

Amendment 1: We planned on conducting all outcome assessments in person with trial participants at 3 and 6 months follow up. We changed that to allow for exceptions to conduct the outcome assessments over the phone when meeting the person was not possible. This was because some trial participants preferred to have the follow-up

assessments over the phone, because of their personal circumstances, or potential for risks involved in meeting the researcher in person (e.g., family restrictions).

Amendment 2a: One of our eligibility criteria for participants was if they could confirm that would be visiting the health facility for the next 3 months. This allowed us to ensure that they would be able to attend the in-person counselling treatment at the health facility, which on an average took around 6-8 weeks to complete. We removed this eligibility criterion as 38% otherwise eligible participants were unable to enter treatment due to this criterion. Additionally, since most participants were not completing the full complement of six sessions, this criterion becomes superfluous. The long time period was also perceived by participants as a commitment that they may not be able to fulfil, deterring them to potentially benefit from the counselling sessions. Finally, we realised from experience that people are unable to predict where they will be residing over the next three months.

Thus, the final eligibility criteria for the trial were as follows:

For the Contact Coverage outcome, health centre attenders were eligible if they are:

• Adults (≥ 18 years);

• Residing in the clusters included in the trial; and

• Speak English or one of the local languages (Konkani, Marathi, Hindi).

In addition to these criteria, health centre attenders were eligible for the Effectiveness outcome if they screened positive for moderately severe or severe depression (total score ≥ 15) on the Patient Health Questionnaire-9 items (PHQ-9).

For both outcomes, we excluded potential participants who meet one or more of the following criteria:

• Patients with significant speech, hearing, or language impairment that interfered with completion of the screening and/or receipt of psychosocial intervention

• Patients who presented to the health centre for emergency medical attention

• Patients with active psychotic symptoms.

Amendment 2b: The IMPRESS trial started on 1st November 2023. Over November and December 2023, the recruitment has been slow, and it was decided at the Annual Review Meeting of the Trial Steering Committee that we would revisit the sample size calculations to ensure feasibility of trial completion. Since the number of clusters had not changed, even with the reduced number of participants we were able to achieve an acceptable power of 80%.

Amendment 3: We revised the sample size further as the actual dropout rate (40%) in the trial was higher than the assumption (20%) that we made in the original sample size. We revised the sample size to be recruited by taking into account a dropout rate of 40% to ensure adequate power to detect the originally hypothesised effect size. The revised power calculations assumed a 40% dropout rate based on our actual experience on this trial, but all other assumptions remained the same.

Amendment 4: As reviewed at the annual meeting of the Trial Steering Committee in January there was an imbalance between the two arms of the trial, both on the numbers screened and the numbers consenting to receive the treatment. This meant that in one of the arms we would have reached the requisite sample size (50% of the total sample size) before the other arm. We proposed to continue screening for participants in both arms until the arm that had slower recruitment also reached the requisite sample size. This effectively means that the arm that was recruiting faster ended up with more participants than originally planned. As we continued to recruit participants using all procedures outlined in the protocol approved by the IRB, this was not a protocol deviation or amendment. Hence, we did not make any change in the protocol.

Amendment 5: The study end date indicated in our annual report was March 2025. We amended that to 31st March 2026.

Amendment 6: We included an additional data collection point for one of the outcomes. This revision was driven by a recommendation made by the Investigator Team to have adequate time points for the trial mediators.

The depression awareness questionnaire is a bespoke tool to assess community awareness about depression. We originally planned to administer this tool at baseline, 6

and 12 months of implementation. The questionnaire underwent minor modifications during the course of the trial, and hence we administered this tool at 18 months of implementation, to ensure that we have the necessary time points considered for mediation analysis. An additional wave of data would allow the predictor, mediator, and outcomes to be temporally sequenced at different time points.

2.3 Trial Arms

Facility Model: The HAP, a manualized and evidence-based psychosocial treatment based on behavioural activation, will be delivered by the trained counsellors. HAP includes the following strategies: psychoeducation, behavioural assessment, activity monitoring, activity structuring and scheduling, activation of social networks, and problem-solving. Homework is assigned between sessions. HAP will be delivered in an individual format. It entails three phases of treatment, delivered over a maximum of six to eight weekly sessions, each lasting up to 40 minutes. Sessions will be delivered face-to-face, at the health centre where the counsellors are already employed.

The HAP intervention was delivered by existing health care workers from the participating health centres. These individuals (called 'counsellors' from here onwards) were nominated for the HAP training by the Health Officers managing the health centres. HAP Clinical Facilitators are experienced project-employed staff who have acquired relevant clinical competencies in delivering HAP through training and supervised HAP delivery. They will supervise HAP Counsellors throughout the trial. They will also provide HAP in the health centres as back-up if there are sustained gaps in service in the health centres due to unavailability of the trained counsellors or back-up counsellors. Back-up counsellors are project-employed staff; they will be trained lay-counsellors who will be responsible for providing HAP in the health centres in situations where there is a sustained gap in service due to unavailability of the counsellor from the health centre. This will be a backup arrangement for gaps in service occurring due to a variety of reasons such as transfer of the counsellor or when a counsellor goes on leave.

Community Model: In addition to the Facility Model, village clusters in this arm will receive a community intervention delivered by the *Sangathis* to (i) enhance demand for the HAP

treatment, and (ii) promote engagement with, and completion of, the HAP treatment. The community intervention is co-produced through a rigorous intervention development process, with local community members and includes strategies such as activities to increase awareness about depression (e.g., organising community meetings and street plays), and dissemination of psycho-educational materials (e.g., leaflets and posters), identification of people with possible depression in the community, and facilitation of access to HAP (e.g. accompanying them to the health centre for counselling sessions). Additionally, the *Sangathis* will coordinate continuing care of people receiving HAP, through home visits to encourage behavioural activation, homework completion and following up with the patient, and engaging family members to support the patient in achieving treatment goals.

The community intervention was delivered by *Sangathis* (which means 'a companion' in the local language) who are community members who are respected in their community and have had previous experience working on socio-developmental, livelihood or governmental community programmes.

2.4 Sample size

Trial clusters

During the trial design phase, we realised that existing government-defined catchments allocated to each health centre did not accurately reflect actual health-seeking behaviours and service use patterns. Specifically, people accessed health centres based on their convenience, even if it was in a catchment area which was not their own. Hence health centres could not reliably serve as discrete clusters that could be randomised. Without clearly defined catchments for the health centres, it was difficult to determine which geographical areas could be linked exclusively to specific health centres, an essential requirement for implementing and monitoring the community intervention.

So, our a priori definition of the unit of randomisation was a cluster of villages. We used a Geographic Information System (GIS)[12] tool to assess the internal validity of our definition by mapping villages to the health centres in which their residents accessed

healthcare. To reduce the risk of contamination between the two arms, a village was included in a particular cluster if (a) 75% of the healthcare-seeking population from that village accesses services at the health centre associated with that cluster, and (b) it had contiguous borders with other villages in the cluster. This resulted in 28 discrete clusters that we randomised.

Contact coverage outcome

Original power calculation: We screened for five days a week for 12 months (52 weeks) in all 30 health centres. Based on screening numbers over the first two months of the trial (average of 6624 per month), we will be able to screen 79488 over 12 months. Based on the screening data from the first two months of the IMPRESS trial, the prevalence of depression (PHQ-9 score >4) among those screened at the health centres is 14%. Assuming this prevalence in health centres in the Community Model this sample size provides 86% power to detect a difference with a prevalence of 11% in those screened at the health centres in the Facility Model at the 5% significance level, assuming a coefficient of variation of 0.2 as below.

Effectiveness outcome

Original: 21 participants with a PHQ-9 score $\geq$15 will be recruited from each of the 28 clusters (N=588). Allowing for 20% of participants to be either not traceable by the outcome assessors or to not consent to participate at follow-up, we expect 17 participants from each cluster to complete the outcome evaluation at the three-month endpoint. This gives a sample size of 476 in which to assess our effectiveness outcome at three months. This sample size provides 80% power to detect an effect size (standardised mean difference (SMD)) of 0.4 at the 5% significance level, assuming a coefficient of variation between clusters of 0.2. The assumptions underlying this calculation are based on the HAP trial for the effect size (i.e. a difference between the mean PHQ-9 scores of 11.9 [SD=7.0] in the control arm and 8.9 [SD=7.2] in the intervention arm) and the MANAS trial for the coefficient of variation (this trial tested a psychosocial intervention for common mental disorders in the same setting)[13,14].

Final: While the trial was underway, we identified that the outcome evaluation dropout rate was higher (40%) than we originally assumed (20%). If we had gone on to recruit 588 as per the original sample size calculation, then with a 40% dropout rate we would have 76% to detect an effect size of 0.4 as originally hypothesised. Hence, we revised our sample size calculations to ensure adequate power to detect the originally hypothesised effect size. The revised power calculations below assume a 40% dropout rate based on our actual experience of the trial, but all other assumptions remained the same.

28 participants with a PHQ-9 score $\geq$15 will be recruited from each of the 28 clusters (N=784). Allowing for 40% of participants to be either not traceable by the outcome assessors or to not consent to participate at follow-up, we expect 17 participants from each cluster to complete the outcome evaluation at the three-month endpoint. This gives a sample size of 476 in which to assess our effectiveness outcome at three months. This sample size provides 80% power to detect an effect size (standardised mean difference (SMD)) of 0.4 at the 5% significance level, assuming a coefficient of variation between clusters of 0.2. The assumptions underlying this calculation are based on the HAP trial for the effect size (i.e. a difference between the mean PHQ-9 scores of 11.9 [SD=7.0] in the control arm and 8.9 [SD=7.2] in the intervention arm) and the MANAS trial for the coefficient of variation (this trial tested a psychosocial intervention for common mental disorders in the same setting[13,14].

2.5 Randomisation

The allocation sequence was generated through restricted randomisation using the following parameters: (a) type of health centre (PHC, CHC, SDH) associated with the cluster; (b) availability of DMHP services in the health centre; and (c) population size of the cluster. A randomly selected seed number was used to enable the randomisation procedure to be reproduced. The randomisation was done before the start of the trial as both the embedding of the trial and other implementation processes (e.g., training of community intervention delivery agents) had to be initiated in advance of recruitment of the first participant.

2.6 Masking

The outcome assessment team was separate from the rest of the team and was blinded to arm allocation. The principal investigators (PIs) and co-investigators, including the trial statistician remained masked until analysis of all primary and secondary outcomes was complete. Given the nature of the interventions and the roles, the following could not be masked: participants, counsellors, *Sangathis*, Case Managers, HAP Facilitators, *Sangathi* Facilitators, data operators and data manager, and clinical coordinators. All outcome data will be analysed masked to the trial arm.

2.7 Timing of outcome assessments and analysis

The contact coverage outcome was measured from 1st November 2023 to 15th July 2025. The effectiveness outcome at three-months (primary outcome) and other secondary outcomes at three-months were measured from 10th February 2024 to 18th November 2025. All secondary outcomes at six-months were collected from 22nd May 2024 to 9th February 2026. The implementation outcomes were collected at different time points as described in the protocol during the implementation of the program. The final analyses for all outcomes will be performed after the SAP is finalized and the dataset is locked.

3 Trial population

3.1 Eligibility criteria

Inclusion

For the Contact Coverage outcome, health centre attenders were eligible if they were:

- Adults (≥18 years);
- Residing in the clusters included in the trial; and
- Able to speak English or one of the local languages (Konkani, Marathi, Hindi).

In addition to these criteria, health centre attenders were eligible for the Effectiveness outcome if they:

- Screened positive for moderately severe or severe depression (total score ≥15) on the Patient Health Questionnaire-9 items (PHQ-9)[15].

Exclusion

For both outcomes, we excluded potential participants (and referred them to existing services if needed) who met one or more of the following criteria:

- Patients with significant speech, hearing, or language impairment that interferes with completion of the screening and/or receipt of psychosocial intervention.
- Patients who presented to the health centre for emergency medical attention.
- Patients with active psychotic symptoms.

3.2 Participant recruitment

A CONSORT flowchart will show recruitment and follow-up of clusters and individual participants.

3.3 Participant recruitment at baseline

Participants were identified and recruited by trained and supervised Case Managers at the health centres. Health centre attenders meeting the eligibility criteria were screened using the PHQ-9 questionnaire[15], a nine-item questionnaire that has been validated and extensively used in the study setting[16,17]. Participants who scored ≥5 (mild depression and above) contributed to the contact coverage outcome.

Due to the limited caseload capacity of the counsellors, we prioritised HAP for the most severely unwell, who have also been shown to derive the greatest benefit from such interventions[18]. This means that only those who scored ≥15 (moderately severe to severe depression) were eligible to receive the HAP intervention and contributed to the effectiveness outcome.

3.4 Participant follow-up

For the effectiveness component of the trial, the endline questionnaire was administered at three- and six-months post-recruitment.

Participants will be categorised as lost to follow up (LTFU) if the outcome assessor failed to meet the participant after making three attempts to either call/ meet the participant on three different occasions (i.e. on three different days at three different points of time) over three weeks (i.e. one attempt made per week). We will describe the proportion of participants who were LTFU at three- and six-months by arm and reasons for the LTFU. We will also identify the baseline factors (e.g. sociodemographic factors, severity of depression) associated with LTFU by comparing those LTFU with those who complete outcome evaluation, separately for 3-months and 6-months endpoints. Participants could withdraw from participation in the intervention, from further data collection, or from all aspects of the study including the use of previously collected data. Data from participants who withdraw consent will only be used as indicated on their withdrawal form. The number and proportion of LTFU and withdrawals will be shown in the CONSORT flow chart.

3.5 Analysis populations

Primary analysis population: For the contact coverage outcome the primary analysis population will be all adults who resided in the randomised clusters. For the effectiveness outcome all enrolled participants (scored ≥15 on the PHQ-9) with endline outcome data will form the primary analysis population. For the effectiveness outcomes, the primary analysis will be conducted according to intention-to-treat (ITT) principles. Participants will be analysed according to the arm they were randomised to regardless of whether they received the intervention.

4 Outcomes

4.1 Data sources

All the data was self-reported and collected through a questionnaire administered by trained research workers. For the contact coverage outcome, the data were collected throughout the screening period. For the effectiveness outcome, data were collected at baseline, three-months post-recruitment (primary endpoint) and six-months post-recruitment.

Further details about data collection and management are given in the trial protocol[19].

4.2 Primary outcome definitions

Contact coverage

Eligible attenders (i.e., residing in one of the 28 clusters) at 30 health centres were screened using the PHQ. Regardless of the health centre they attended, attenders will be allocated to the trial intervention or control arms based on their village/area of residence.

Effectiveness

Eligible attenders in a particular health centre were consented only if they resided in the cluster with which that health centre was associated (see above for explanation). This was critical as the community intervention involved care delivered collaboratively by the counsellor and *Sangathi*.

Table 1 below shows the primary outcomes, their measures, variable types, sources of the data, and the analysis population.

Table 1. Primary outcomes

Primary outcomes	**Measure**	**Variable type**	**Eligibility**	**Data source**
Contact coverage[1]	PHQ-9 score $\geq$5 as a proportion of the cluster population[2]	Categorical	All resident adults in the 28 village clusters	Screening data
Effectiveness	Mean PHQ-9 score	Continuous	All individuals with a score $\geq$15 on the PHQ-9 and consent to receive HAP	3 months questionnaire

[1]For the purpose of our trial we have not described contact coverage as conventionally defined.

[2]In the protocol we had defined Contact Coverage as 'PHQ-9 score >4 as a proportion of those screened'. A key rationale for revising the definition of contact coverage is that the

original denominator (“those screened”) is substantially influenced by operational and contextual factors within health facilities that are not directly related to the intervention. In practice, the number of individuals screened varies due to factors such as fluctuations in outpatient attendance. As a result, the original denominator reflects differences in clinic-level operational performance and hence the original metric risks confounding the intervention’s effectiveness in increasing help-seeking. Using the total cluster population as the denominator provides a more stable and policy-relevant estimate of coverage by anchoring the outcome to the population served by each cluster rather than to variable clinic processes. Importantly, this change does not alter the definition of the numerator but rather improves the validity and interpretability of the coverage outcome as a measure of population reach.

4.3 Secondary outcome definitions

Table 2 below shows the secondary outcomes, their measures, variable types and sources of the data.

Table 2. Secondary outcomes

Secondary outcomes	Measure	Variable type	Eligibility	Data source
Effectiveness at 6 months	Mean PHQ-9 score	Continuous	All individuals who consent to receive HAP	6 months questionnaire
Effectiveness over the 6-month period	Mean PHQ-9 score	Continuous	All individuals who consent to receive HAP	3- and 6- months questionnaire
Remission at 3-months	PHQ-9 score <10	Categorical	All individuals who consent to receive HAP	3- months questionnaire
Remission at 6-months	PHQ-9 score <10	Categorical	All individuals who consent to receive HAP	6- months questionnaire
Recovery	PHQ-9 score <10	Categorical	All individuals who consent to receive HAP	3- and 6-months questionnaire
Response to treatment	>50% reduction in PHQ-9 score	Categorical	All individuals who consent to receive HAP	3- months questionnaire
Sustained response to treatment	>50% reduction in PHQ-9 score	Categorical	All individuals who consent to receive HAP	6- months questionnaire
Depression awareness	Mean depression awareness score	Continuous	Adult health centre attenders	Depression Awareness Questionnaire Baseline at baseline, 6, 12 and 18 months of implementation

Perceived social support	Mean MSPSS score Mean bespoke tool score	Continuous	Individuals who consent to receive HAP	3- and 6-months questionnaire
Treatment completion	% participants who met treatment goals or completed the maximum number of sessions or were referred to mental health specialists	Categorical	Individuals who receive HAP	PEERS data
Behavioural activation	Mean PREMIUM Abbreviated Activation Scale score	Continuous	Individuals who consent to receive HAP	3- and 6-months post-recruitment
Costs	Inventory form for collecting system-level economic costs of delivering interventions	Continuous	Individuals who consent to receive HAP	3- and 6-months post-recruitment
Costs	Client Service Receipt Inventory form for collecting costs to patients	Continuous	Individuals who consent to receive HAP	3- and 6-months post-recruitment

Costs	Survey form for collecting costs of receiving HAP intervention to patients	Continuous	Individuals who consent to receive HAP	3- and 6-months post-recruitment
Disability	Standardized disability scores used to estimate QALYs using WHODAS	Continuous	Individuals who consent to receive HAP	3- and 6-months post-recruitment

5 Statistical analyses

5.1 Baseline characteristics and comparability by arm

Descriptive statistics for health centre- and individual-level baseline characteristics will be shown by arm and will include the mean and standard deviation (SD) (or median and interquartile range if non-normal distribution) for continuous variables and frequencies and percentages for categorical variables.

The following characteristics will be described:

For analysis of contact coverage:

- **Cluster-level:** Population size, participants positive on PHQ-2.

For analysis of effectiveness:

- **Cluster-level:** Participants scoring ≥15 on PHQ-9, participants eligible for consenting, and participants consented.
- **Individual level:** Age, gender, marital status, education, employment status, WHODAS 2.0, MSPSS, PAAS, PHQ-9.

Baseline imbalance by arm will be explored. We will identify substantial differences between arms in terms of the individual level factors and will adjust for these differences

in adjusted effectiveness outcome analyses. No formal statistical testing will be performed to examine differences in baseline characteristics between the trial arms, as any difference would have been due to chance if the randomisation was correctly performed. We will finalise the factors for adjusting before the effectiveness analysis is conducted.

For the outcomes we will plot histograms within each arm to assess for normal distribution to determine how to describe the outcome. Transformations or use of robust SE will be carried out if the continuous data are not normally distributed. Continuous data that are approximately normally distributed will be summarised in terms of the mean, standard deviation, median, minimum, maximum and number of observations. If a transformation is not possible, data will be presented in terms of the maximum, median, minimum and number of observations. Categorical data will be summarised in terms of frequency counts and percentages.

The analyses as described below will be at the individual level (primary) and cluster level (secondary).

5.2 Intra-cluster Correlation Coefficient (ICC) and coefficient of variation

The ICC and coefficient of variation will be estimated and reported based on primary and selected secondary outcomes.

5.3 Analysis of continuous outcomes

Effectiveness

Random effects linear regression will be used to compare mean scores for continuous outcomes by arm, adjusting for clustering, baseline cluster-level severity of depression, baseline individual-level of the outcome variable, factors used for restricted randomization, and any confounding factors (provided there is no collinearity between individual factors) for which there is substantial imbalance between the arms at baseline. We will present an estimate of the effect size as an absolute mean difference (AMD) adjusted standardized mean difference (SMD) and 95% confidence intervals, and the

intra-cluster correlation. In case of problems with convergence, we will use robust standard errors.

5.4 Analysis of binary outcomes

Contact coverage

The primary analyses will compare the proportion of individuals identified with depression through screening in the health centres. The risk ratio will be estimated using binomial regression for the cluster-level data will be used to compare the number of individuals between arms, with the cluster population size as the denominator.

Effectiveness

For binary outcomes, the effect will be measured by the risk ratio (RR) and 95% confidence intervals using modified Poisson regression models adjusted for clustering, baseline severity of depression, factors used for restricted randomization, and any confounding factors for which there is substantial imbalance between the arms at baseline. In case of problems with convergence, we will use robust standard errors, or present odds ratios with their 95%CI.

Of note, the estimates from random effects models are interpretable as the effect of treatment on an individual in the median cluster, not the population-level impact of treatment across all clusters[20]. The estimate generated using a random-effects model will thus be useful in understanding the clinical impact of the community intervention on an individual patient but will not fully reflect the public health impact of implementing across all clusters. If the results of secondary outcomes analysis indicate that the community intervention has a moderate ($p<0.1$) impact on binary outcomes (i.e., remission, response to treatment), we will conduct a sensitivity analysis using population average models (xtgee command in Stata). However, the results from fully adjusted random effects models will represent the primary analysis of binary outcomes.

Interpretation of the intervention effect will be based on the strength of evidence of effect size and consistency of results for related outcomes.

5.5 Analysis of outcomes with repeated measures per participant

Outcomes measured at 3- and 6-months for each participant will be analysed using a mixed-model for repeated measures (MMRM) that accounts for both clustering of participants within clusters and within-participant correlation across time. The model will include a random intercept for cluster and an appropriate covariance structure for repeated measures, with fixed effects for time, trial arm, and their interaction, and adjustment for prespecified baseline covariates. In case of problems with convergence, we will use robust standard errors, or present odds ratios with their 95%CI.

5.6 Effect-modification and mediation

Pre-specified potential effect modifiers are given in the table below.

Effect-modification for each of these variables with trial arm will be assessed for primary and selected secondary outcomes (effectiveness, effectiveness over the six-month period, remission, recovery, response to treatment, sustained response to treatment) by fitting an interaction term with each variable and trial arm in the model.

List of potential effect modifiers

Effect modifier (baseline values)	**Measure**
Contact coverage	
Size of health centre	Binary (PHC vs CHC/SDH)
Screening period	Quarters (three-month periods) for the duration of screening
Effectiveness	
Age group	Binary (categories to be decided based on distribution of data and agreed before unblinding)
Gender	Binary
Marital status	Binary (married/in a relationship vs single/divorced/separated/widow(er))
Education status	Binary (categories to be decided based on distribution of data and agreed before unblinding)

Employment status	Binary (categories to be decided based on distribution of data and agreed before unblinding)
Severity of depression (at baseline)	Binary (moderately severe vs severe)
HAP implementation period	Quarters (three-month periods) for the duration of HAP delivery

We will also explore hypothesised a priori causal pathways through mediation analyses.

Mediator	**Measure**	**Variable type**
Contact coverage		
Community awareness	Depression Awareness Questionnaire	Continuous
Effectiveness		
Perceived social support (generic)	Multidimensional Scale of Perceived Social Support (MSPSS)	Continuous
Perceived social support (intervention specific)	Bespoke tool	Continuous
Treatment completion	Process data	Binary
Behavioural activation	PREMIUM Abbreviated Activation Scale (PAAS)	Continuous

Using MPlus, we will conduct causal mediation analysis using the Monte Carlo Method for Assessing Mediation (MCMAM) to assess the effects of a priori and theoretically-driven mediators on primary outcomes[21]. MCMAM performs better than the Sobel test and comparably with bootstrapping approaches[22]. Using structured equation modelling (SEM), we will conduct a multiple mediation model examining all significant mediation pathways[23]. All analyses will adjust for relevant baseline variables as well as potential covariates, including patient-, counsellor- and treatment-level characteristics as appropriate[24].

5.7 Sensitivity and additional analyses

The following analyses will be conducted for primary and selected secondary outcomes (effectiveness, sustained effectiveness, remission, sustained remission, response to treatment, sustained response to treatment), and the results will be compared.

Sensitivity analyses:

- We will do a sensitivity analysis to examine the contact coverage for moderate depression and above (i.e., PHQ9≥10)
- We have estimated contamination at baseline by asking those who are screened whether they have been exposed to any of the components of the community intervention (e.g., attended awareness events, saw flyers). These data will be used to conduct sensitivity analysis estimating the contamination range and the actual contact coverage effect after excluding those in the control arm who report any exposure to the community intervention.
- We will examine effectiveness after excluding participants who had no contact with their assigned community volunteer.
- We will examine effectiveness after excluding participants for whom outcomes were assessed outside of 3- and 6- months window

Additional analyses:

- Process measures measured within the trial arms will be summarized as means and proportions as appropriate; and with 95% confidence intervals.
- We will estimate complier average causal effects based on treatment completion.
- Compare between the two arms, contact coverage as a proportion with PHQ-9 score of ≥5 amongst screened participants who belong to the trial clusters.
- Assess the effectiveness separately for and compare between the counsellors (health system staff) and HAP Facilitators/Back up counsellors (programmatic staff).

5.8 Exploratory analyses

We will undertake exploratory moderator analyses as follows:

1) Whether the type of health system counsellor moderates the effect of the intervention on the primary outcome. This includes analysis AYUSH vs non-AYUSH practitioners.

5.9 Economic evaluation

5.8.1 Cost-effectiveness analysis

We will compare the costs and effectiveness between the Facility and Community Models using the incremental cost-effectiveness ratio (ICER). To generate incremental costs between the models, we will use multi-level generalized linear models (and controlling for clustering effects for skewed cost data) at the individual patient level. Costs per patient will be the outcome variable. A dichotomous variable indicating the Community or Facility models is an exposure variable, with Facility model as the reference. Other covariates include time and interaction terms between intervention dummies and time. To provide means and 95% confidence intervals (CIs) for the estimated incremental costs, we will use bootstrapping methods. The same strategy can be applied to incremental effectiveness and ICER for generating their means and 95% CIs. We will generate cost-effectiveness acceptability curves with different levels of thresholds on willingness to pay, from either a health system perspective or a societal perspective, to show the probability that the intervention remains cost-effective as thresholds change.

5.8.2 Long-term sustainability modelling

If the intervention is cost-effective, we will further assess the scaling-up costs of the Community Model (which includes the HAP) and propose recommendations on its long-term financial sustainability.

The sustainability analysis will focus on whether the HAP plus community (Sangathi) treatment component of the trial can be realistically implemented at scale within the public health system, moving beyond trial conditions. We will combine evidence from the trial with information on how the health system operates in practice, including staffing, budgets, and infrastructure, to understand what it would take to expand this intervention

beyond the study setting. In particular, we will estimate the resources required to scale across districts or states and assess how this compares with existing allocations under programmes such as the District Mental Health Programme (DMHP). An important part of this exercise will be to identify where the main constraints lie—whether in financing, workforce capacity, or the ability to generate and sustain demand for care.

Our approach will bring together four components. First, we will conduct a cost-effectiveness analysis to estimate the cost per treated case and per remission under realistic implementation assumptions. Second, we will develop a budget impact model that scales these costs using data on population size, prevalence, and service capacity to generate district and state level estimates. Third, we will use a coverage-based framework that follows individuals from need to treatment initiation and completion, allowing us to identify where drop-offs occur and to quantify the extent to which the community component of the RCT improves uptake and adherence. Finally, we will also explore uncertainty in key parameters through sensitivity analyses to assess how results vary under different assumptions about costs, uptake, and implementation quality. The aim is to produce transparent and policy-relevant estimates that can inform decisions around scaling the intervention within the public system.

6 Statistical considerations

6.1 Missing data

Missing data on outcomes and key covariates will be assessed prior to analysis. In situations where >5% of data are missing, we will use missing data methods appropriate for random effects models[25]. Data will be assumed to be missing at random (MAR), and respondents with missing data will be described by key sociodemographic and clinical characteristics. We will calculate multiple imputation (MI) models using linear or logistic regression models as appropriate. We will adjust for factors associated with missingness as fixed effects in imputation models, and 50 datasets will be imputed. Where MI models are used, these will be the primary analysis results, provided they are able to be estimated with no difficulty. Complete case analysis will be conducted and reported as a sensitivity analysis and will be used as the primary analysis if there are difficulties obtaining estimates in the MI models.

6.2 Multiple testing adjustment

The two primary outcomes are considered independent due to different causal mechanisms, so an improvement in one without the other will be informative.

For the secondary outcomes, specific inferences will be made for each individual null hypothesis, and multiple testing adjustment will not be performed. Given the number of tests performed, the intervention effects will be interpreted with caution[26,27].

7 AEs and SAEs reporting

Descriptive statistics including frequencies and percentages will be used to summarise and present AEs and SAEs by trial arm. Where appropriate, the results will be stratified by clusters.

8 Implementation Evaluation

Implementation indicators will be analysed primarily descriptively and will be used to characterise intervention delivery, fidelity, feasibility, acceptability, and sustainability.

Adoption will be assessed at the counsellor, Sangathi and health-system levels. Quantitative adoption indicators will be summarised using frequencies and proportions for categorical variables and means (standard deviations) or medians (interquartile ranges) for continuous variables, as appropriate. ORIC scores will be summarised by trial arm and time point (baseline and 12 months). Qualitative data from IDIs and FGDs will be analysed using thematic analysis, with findings mapped to Adoption constructs and triangulated with quantitative indicators.

Implementation indicators will assessed at the patient, counsellor, Sangath and health systems level. Binary indicators will be summarised as proportions with 95% confidence intervals. Continuous indicators will be summarised using means (standard deviations) or medians (interquartile ranges), as appropriate. For indicators measured before and after (e.g., knowledge and competency scores), within-person changes will be assessed using paired analyses and summarised as mean differences with 95% confidence intervals. Where implementation indicators are available at the cluster level, summaries will be presented by trial arm.

Maintenance outcomes will be derived primarily from qualitative interviews with counsellors, Sangathis, health-system stakeholders and decision-makers, focusing on perceived sustainability, continued programme delivery, organisational readiness, and barriers and facilitators to long-term implementation. Qualitative data will be analysed using thematic analysis and integrated with quantitative implementation indicators to provide an overall assessment of programme sustainability.

Process and fidelity indicators will be analysed to characterise intervention delivery and to aid interpretation of the effectiveness findings. These analyses are intended to describe implementation performance rather than formally test intervention effectiveness. Process indicators described in the protocol will be summarised descriptively.

9 Concomitant Care

Concomitant mental health care reported during follow up will be summarised descriptively by trial arm.

10 Adverse Events and Serious Adverse Events

Following summaries of serious adverse events (SAEs) reported during the trial will be presented by arm: Number of participants experiencing at least one SAE, total number SAEs, categories of SAE, relatedness to intervention. Given that the trial is not powered for safety outcomes, all AE and SAE analyses will be considered descriptive and exploratory.

11 Software

All analyses will be performed using STATA Version 18 or 19 and MPlus.

12 References

1. Collaborators GMD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry.* 2022;9(2):137-150.
2. Santomauro DF, Herrera AMM, Shadid J, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet.* 2021;398(10312):1700-1712.
3. Sagar R, Dandona R, Gururaj G, et al. The burden of mental disorders across the states of India: the Global Burden of Disease Study 1990–2017. *The Lancet Psychiatry.* 2020;7(2):148-161.
4. Moitra M, Santomauro D, Collins PY, et al. The global gap in treatment coverage for major depressive disorder in 84 countries from 2000–2019: A systematic review and Bayesian meta-regression analysis. *PLoS Medicine.* 2022;19(2):e1003901.
5. Gururaj G, Varghese M, Benegal V, et al. National Mental Health Survey of India, 2015-16: Summary. *Bengaluru, National Institute of Mental Health and Neuro Sciences, NIMHANS Publication.* 2016(128).
6. van Ginneken N, Chin WY, Lim YC, et al. Primary-level worker interventions for the care of people living with mental disorders and distress in low- and middle-income countries. *Cochrane Database of Systematic Reviews.* 2021 2021(8).
7. Singla DR, Kohrt BA, Murray LK, Anand A, Chorpita BF, Patel V. Psychological treatments for the world: lessons from low-and middle-income countries. *Annual Review of Clinical Psychology.* 2017;13:149-181.
8. Sarikhani Y, Bastani P, Rafiee M, Kavosi Z, Ravangard R. Key Barriers to the Provision and Utilization of Mental Health Services in Low-and Middle-Income Countries: A Scope Study. *Community Ment Health J.* Jul 2021;57(5):836-852.
9. Chowdhary N, Anand A, Dimidjian S, et al. The Healthy Activity Program lay counsellor delivered treatment for severe depression in India: systematic development and randomised evaluation. *The British journal of psychiatry.* 2015:bjp. bp. 114.161075.
10. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical care.* 2012;50(3):217.
11. Gaglio B, Shoup JA, Glasgow R. The RE-AIM framework: a systematic review of use over time. *American Journal of Public Health.* 2013;103(6):e38-e46.
12. Nykiforuk CI, Flaman LM. Geographic information systems (GIS) for health promotion and public health: a review. *Health promotion practice.* 2011;12(1):63-73.
13. Patel VH, Kirkwood BR, Pednekar S, et al. Improving the outcomes of primary care attenders with common mental disorders in developing countries: a cluster randomized controlled trial of a collaborative stepped care intervention in Goa, India. *Trials.* 2008;9:1-11.
14. Patel V, Weiss HA, Chowdhary N, et al. Effectiveness of an intervention led by lay health counsellors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomised controlled trial. *Lancet.* 2010;376:2086-2095.
15. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals.* 2002;32(9):509-515.
16. Patel V, Weobong B, Weiss HA, et al. The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. *The Lancet.* 2017;389(10065):176-185.
17. Patel V, Araya R, Chowdhary N, et al. Detecting common mental disorders in primary care in India: a comparison of five screening questionnaires. *Psychol Med.* 2008;38:221-228.

18. Driessen E, Cuijpers P, Hollon SD, Dekker JJ. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *Journal of consulting and clinical psychology.* 2010;78(5):668.
19. Nadkarni A, Gandhi Y, Fernandes L, et al. Effectiveness and cost-effectiveness of a community intervention in enhancing access to care and improving clinical outcomes for depression: a protocol for a cluster randomised controlled trial in India. *Trials.* 2024/08/28 2024;25(1):569.
20. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology.* Jul 2010;21(4):467-474.
21. MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research.* 2004;39(1):99-128.
22. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychological methods.* 2002;7(1):83.
23. Valente MJ, MacKinnon DP. Comparing models of change to estimate the mediated effect in the pretest–posttest control group design. *Structural equation modeling: a multidisciplinary journal.* 2017;24(3):428-450.
24. VanderWeele T, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface.* 2009;2:457-468.
25. Drechsler J. Multiple Imputation of Multilevel Missing Data—Rigor Versus Simplicity. *Journal of Educational and Behavioral Statistics.* 2015/02/01 2015;40(1):69-95.
26. Li G, Taljaard M, Van den Heuvel ER, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int J Epidemiol.* Apr 1 2017;46(2):746-755.
27. Parker RA, Weir CJ. Multiple secondary outcome analyses: precise interpretation is important. *Trials.* 2022/01/10 2022;23(1):27.