BMC Medicine

RESEARCH

Open Access

A pilot cluster-controlled trial of interventions to improve detection of depression in primary healthcare in Ethiopia

Abebaw Fekadu[1,2,3*], Barkot Milkias[2], Rahel Birhane[1], Bethlehem Tekola[4], Mahlet Yared[2], Tigist Eshetu[1], Tomas Getahun[1], Heidi Lempp[5], Andre Tylee[4], Merga Belina[1], Girmay Medhin[1,6], Kassahun Habtamu[7], Atalay Alem[2], Crick Lund[4], Inge Petersen[8], Graham Thornicroft[4], Anthony J. Cleare[9], Rosie Mayston[4], Esubalew A. Assefa[1,10], Mekdes Demissie[1,11], Charlotte Hanlon[12] and Martin J. Prince[4]

Abstract

Background The low recognition of depression in primary healthcare (PHC) remains a major obstacle to rendering adequate care for people with depression globally. This study aimed to evaluate the feasibility and potential benefit of a contextually developed multicomponent and multilevel intervention to improve the identification of depression in PHC.

Methods A pilot, four-arm, parallel-group, cluster, non-randomised controlled trial was conducted in a predominantly rural district in Ethiopia. The active interventions were allocated to three PHC facilities: (1) a core multicomponent intervention focusing on providers—a manualised training package along with system intervention (mobile application, posters, quality improvement and supervision) (Level-I/Arm I), (2) Level-I intervention plus a 4-item screening questionnaire administered by triage nurses (Level-II/Arm II), (3) Level-II intervention plus service user awareness raising (Level-III/Arm III). In the control facility, standard integrated mental healthcare (care by providers trained in the standard WHO mhGAP intervention guide) was available. The outcomes were the identification of depression and the feasibility and acceptability of implementation by PHC clinicians. Quantitative and qualitative data were collected post-intervention. Descriptive analysis and thematic analysis were used to analyse the data.

Results A total of 21 providers (14 clinicians and 7 triage nurses) and 1659 adult outpatients participated in the study. Overall, 116 outpatients (7.0%) received a diagnosis of depression by PHC clinicians. Detection of depression was significantly better in the active intervention arms combined: 8.3% ($n = 115/1380$) vs. 0.4% ($n = 1/279$) in the control arm. Level-II and Level-III intervention arms had significantly higher rates of detection (10.1% Level II, 9.2% Level III) compared with Level I (5.2%); however, there was no significant difference between Level-II and Level-III. The interventions demonstrated very good acceptability, feasibility and appropriateness although screening, which was included in the Level II and Level III intervention arms, had relatively lower acceptability and an overall low positive predictive value.

*Correspondence:
Abebaw Fekadu
abebaw.fekadu@aau.edu.et
Full list of author information is available at the end of the article

Conclusions The tested interventions hold good promise of enhancing the identification of depression in PHC, with excellent feasibility and acceptability parameters. Although screening has good potential, implementation in routine care requires further optimisation and evaluation.

Trial registration Trial ID: PACTR202206723109626. Registration date: 21 June 2022. Retrospectively registered.

Keywords Depression, Detection, Primary healthcare, Intervention, Feasibility, Acceptability, Pilot trial, Africa

Background

The global public health relevance of depression is incontrovertible. It is the second most prevalent mental health condition, affecting up to 1 in 5 adults over a lifetime [1–3]. Approximately one third of primary healthcare (PHC) attendees experience impairing symptoms of depression [3] and depression accounts for the largest proportion of mental disorder disability adjusted life years [4]. However, most people with depression, particularly in low and middle-income countries (LMICs), are untreated [5–8] or do not receive minimally adequate treatment [9]. In Ethiopia, there is a strong commitment to scale up mental healthcare through the primary care system [10] and evidence for the effectiveness of this approach for treating people with psychosis [11] but not for depression. While the low treatment rate in LMIC may partly be due to the scarcity of specialists [9], PHC plays a critical role in the care of depression even in settings with advanced health systems [2], with up to 9 in 10 episodes of depression managed in PHC [12]. However, on average half of the cases attending PHC go undetected [3, 13], with detection rates close to 0% in LMIC [3].

Studies from Africa have reported detection rates close to 0%, which do not improve with brief [14, 15] or more intensive training [16]. A more elaborate intervention programme was implemented in Kenya as part of a national initiative supported by the World Health Organization (WHO) [17]. In this programme, primary care practitioners were offered a 5-day structured training, which included teaching on the core concepts of mental disorders and their practical significance, communication, assessment and management skills, and relevant health system issues, such as policy, legislation and comorbidity. The study did not find impact on detection of mental disorders [16]. However, a similar study implemented in Malawi has found significant improvement in detection of depression [18].

In Ethiopia, three initiatives, which have used training with the Mental Health Gap Action Programme Intervention Guide (mhGAP-IG) as a core training tool, have explored the detection of depression by PHC clinicians. The pioneering collaborative work of the Ethiopian Ministry of Health and the WHO described the number of 'cases detected' in 19 health facilities over a 6-month period. Just 89 of 592 (15%) service users had received a diagnosis of depression. Another study supplemented the mhGAP training with 5 days of practical training. This study reported low rate of detection at the beginning [19], which did not change after programme implementation. A larger-scale pre- and post-intervention study that trained 94 health professionals evaluated the impact of the training intervention after 3 months of practice [20]. The rate of diagnosis increased from 15.87% ($n=10/63$) to 18.75% ($n=54/288$), an increase of under 3%.

The lack of improvement in the detection of depression from training is not surprising [21, 22]. Improvements in detection require more complex approaches [12, 21, 22]. This has been confirmed by a recent systematic review of the global evidence [23] initially reviewed to April 2020 and updated to April 2024 as part of this report. The evaluated interventions included clinician training, implementation of guidelines, collaborative care packages or quality improvement (QI) programmes, screening, screening with feedback, a combination of training, screening and feedback, using request of antidepressant prescription as a prompt and an eclectic intervention consisting of training, leadership enhancement, dissemination, support and auditing. Trainings that focused on active learning, role play and clinical practice were linked to improvements in the detection of depression; however, effects were typically not sustained for longer than 6 months. Screening with feedback was more effective in improving the detection of depression than screening alone. In most of the studies that used clinician training in addition to screening and feedback, improved recognition of depression was reported. Additionally, most studies reported that implementation of guidelines/collaborative care packages/quality improvement programmes increased the detection of depression in the PHC setting. The evidence was primarily from high-income countries and about three-quarters of the included studies were methodologically weak or moderate.

To address this critical evidence gap, we iteratively co-developed multicomponent and multilevel interventions [24], applying the principles of the Medical Research Council's (MRC) framework for the development and evaluation of complex interventions [25] (Additional file 1: Fig. S1), supplemented by the Bowen's Feasibility Framework [26].

The objective of the current pilot study was to evaluate the potential utility of the new multicomponent, multilevel intervention to improve detection of depression and whether the process of implementation is perceived as feasible, acceptable and appropriate.

Methods

Study design and setting

The primary study design was a pilot, four-arm, cluster, non-randomised controlled trial (PACTR202206723109626). The process of implementation was evaluated following the MRC framework [25] and Bowen's Feasibility Framework [26] (Additional file 1: Table S1) through (1) cross-sectional quantitative assessment of feasibility, acceptability and appropriateness of the intervention components and (2) embedded qualitative in-depth interviews and focus group discussions of feasibility, acceptability and perceived utility of the intervention packages. The impact of the intervention packages in improving detection of depression by PHC practitioners was evaluated through extraction of diagnosis recorded in the clinical records of service users.

The study was conducted in the Sodo and south Sodo districts, and Bui town administration of the East Gurage Zone, of the Central Ethiopia Regional State, located about 100 km south of the capital city, Addis Ababa (Fig. 1). These districts host a UNESCO world heritage site and are served by one primary hospital, eight health centres and over 50 health posts. The primary hospital is used by the population of the three districts, while each health centre serves a catchment area population of 9000 to 40,000 residents (Additional file 1: Table S2).

Participants

The study population included clinicians working in outpatient clinics and triage nurses, who administered the interventions, and outpatients aged 18 and above attending selected facilities for the first time during the month of the study (May 2021). Outpatients who had been diagnosed or treated for a depression previously or assessed by non-trained PHC workers were excluded.

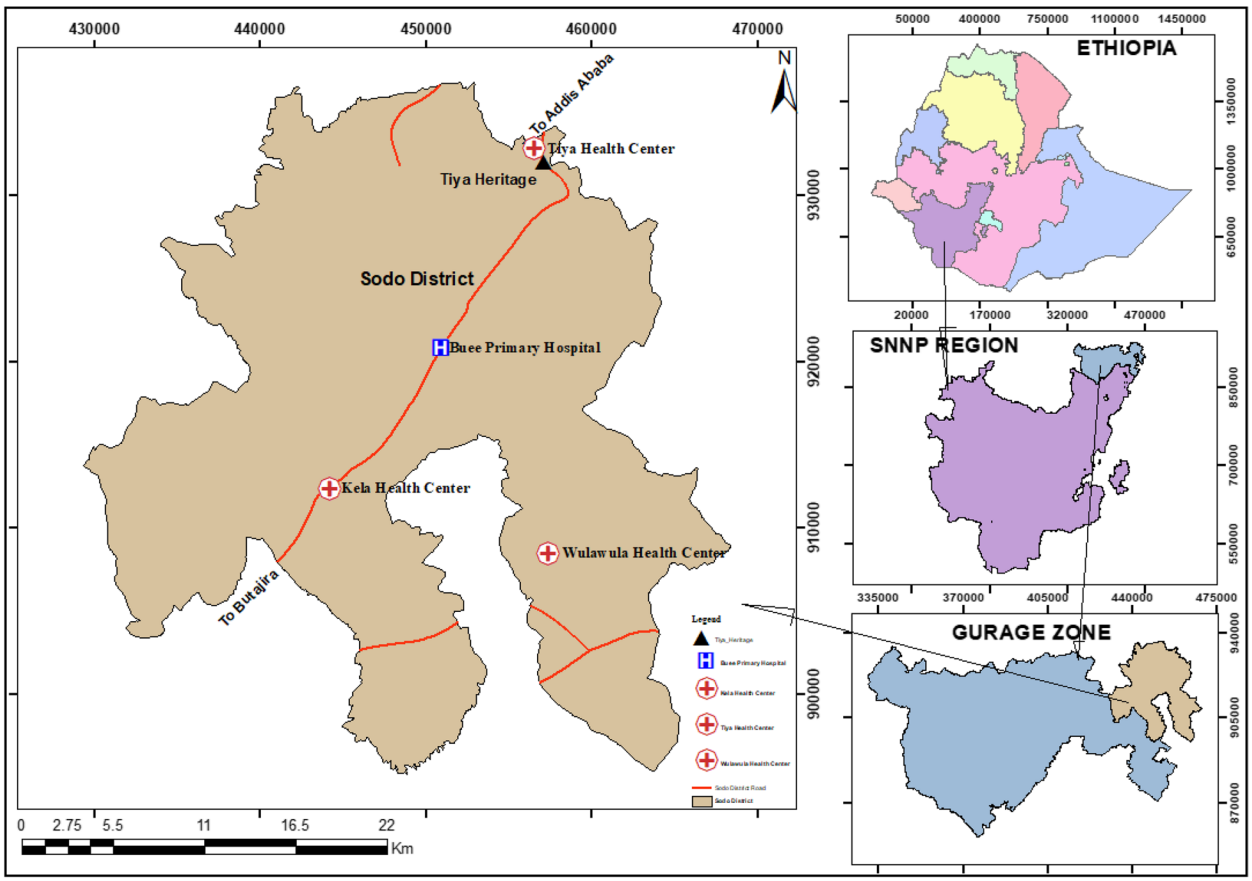

Fig. 1 Map of the study setting -- study districts and health facilities where the study was conducted

Intervention packages

Intervention package development

We co-developed the multicomponent intervention packages [23, 24] by applying the principles of the MRC's framework through four non-sequential steps [25] (Additional file 1: Fig. S1). First, extensive formative research, including systematic reviews, in-depth interviews, and quantitative cross-sectional and cohort studies, was conducted to have a more contextualised understanding of the nature, common presentations and impact of depression [27, 28]. Secondly, we identified and collated promising interventions through a systematic review of the global literature [23]. Thirdly, screening tools were assessed for potential utility and acceptability [29] with a new four-item screening tool adapted for use. Finally, two consensus meetings were convened with mental health experts, primary care clinicians, patients and healthcare administrators to agree on the key intervention components, considering feasibility, acceptability and utility. We have published the main intervention components [24].

Components of the intervention packages

The intervention packages included the following:

i. Manualised training package: Training of PHC clinicians used a manualised training package adapted from the World Health Organization's mental health Gap Action Programme Intervention Guide (mhGAP-IG), with emphasis on depression, active learning methods, communication skills and involvement of people with lived experience as trainers to share their experience of depression treatment pathway. A 5-day theoretical and 5-day practical (on the job) training was facilitated by a psychiatrist.
ii. Screening: This included screening outpatients for culturally salient manifestations of depression, using a four-item tool with feedback to clinicians. The four items consisted of the original 2-items of the PHQ-2 (depressed mood and loss of interest) and irritability and noise intolerance [28, 29]. All four items were rated as simple binary responses—'yes' (present) or 'no' (absent). On average, completion of the screening took 1.8 minutes [24].
iii. Service user awareness raising: Two pamphlets (information leaflets each a page long, describing general mental wellbeing and depression) were prepared to enhance awareness amongst people attending outpatient clinics. This was augmented with direct education provided by experienced field workers in the selected site. This intervention was considered relevant based on the initial formative work in which low awareness of the community about depression was considered an important barrier to the reporting of depressive symptoms by patients and thus the recognition of depression.
iv. System level interventions: As part of a cross-cutting system level intervention, supportive supervision by a psychiatric nurse, quality improvement meetings, posters in outpatient clinics as aide-mémoire and a decision support mobile application were provided.

The manualised training package and the system level interventions constituted the core intervention package.

Intervention and control groups

Three active intervention groups, representing three levels of intervention intensity, were compared against an existing 'integrated' mental healthcare programme, control group (Fig. 2).

Level-I intervention package (Arm 1): The health facility in this arm received the core intervention, consisting of manualised training of PHC clinicians and

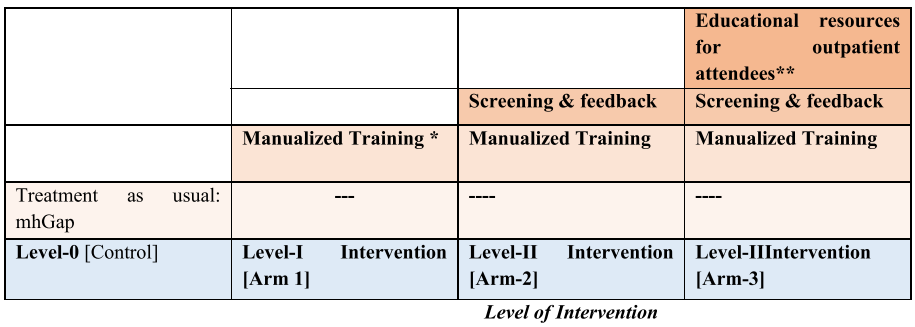

Fig. 2 Intervention package components and levels of intensity representing the three active intervention arms. *System level intervention components: supportive supervision, quality improvement, mobile app, poster and pocket manual. **Information leaflets and direct education of outpatient clinic attendees. mhGAP, mental health Gap Action Programme Intervention Guide

cross-cutting system level intervention package as described above, which was also included for all the active intervention arms. All clinical staff working in the outpatient clinics and providing clinical assessment and treatment were included as providers irrespective of their qualification.

Level-II intervention package (Arm 2): Level-I or core intervention package *plus* administration of the 4-item screening tool, PHQ-2 + 2 [28, 29]. The screening was administered by triage nurses of the selected facilities and recoded in triplicate carbon copy pads with one copy retained within the triage pad, a second attached to the clinical records and the final copy submitted to the research team. The screening was completed for all adult outpatients of the selected facilities throughout the 4 weeks of implementation. Any 'yes' response to any of the four items was considered screen-positive. The screening was passed on to the clinician, who makes further assessment for diagnostic decision. The triage nurses were trained on the use of the screening items, how to handle the questions with sensitivity, and provide the completed screening to the clinician. They were also supervised initially in as part of the training exercise.

Level-III intervention package (Arm 3): Level-II intervention package *plus* service user awareness raising as described above.

Control arm (Level 0): Control arm clinicians received the standard mhGAP-IG-based training to support a broader integrated mental healthcare for priority conditions (psychosis, depression, epilepsy, alcohol use disorder and other common mental disorders).

Selection and allocation of intervention clusters

Selected facilities: Three PHC facilities (Bui Primary Hospital, Kela Health Centre and Wula wula Health Centre) were chosen non-randomly based on the larger number of their catchment area population and one facility for being entirely rural. Control facility was chosen based on the availability of staff trained in the mhGAP-IG to provide mental healthcare.

Intervention allocation: Facilities with higher outpatient flow were allocated to interventions that included screening (Levels II and III). The most urban facility, assumed to have more literate outpatients, was allocated to the Level-III intervention.

Outcomes

The primary outcome was the identification of depression by PHC clinicians. Process outcomes, such as knowledge and competency of the PHC workers, and feasibility, acceptability and appropriateness of interventions were secondary outcomes (Table 1 and Fig. 3). The intervention was implemented immediately after the completion of training. The implementation lasted 1 month during which all the detection data were collected.

Detection of depression

Detection was equated with a recorded diagnosis of depression in the clinical notes of outpatients by the

Table 1 Assessment of outcomes

Evaluation parameter	Outcome assessed	Evaluation method[a]	Assessment time	Interviewer/assessor
Primary outcome	Detection (by PHC clinicians)	Data extraction from clinical records	Continuous over 1-month post-intervention	Trained data collectors
Secondary outcomes	Treatment	Data extraction from clinical records	Post-intervention for 1 month	Psychiatric nurses
	Accuracy of diagnosis	Confirmatory diagnosis	Post-intervention	Psychiatrists using MINI
Process outcomes	Acceptability	Acceptability of Intervention Measure, four items	Immediately post-intervention	Self-administered
	Appropriateness	Intervention Appropriateness Measure, four items	Immediately post-intervention	Self-administered
	Feasibility	Feasibility of Intervention Measure, items	Immediately post-intervention	Self-administered
	Training fidelity	Fidelity and completion checklist	Daily during training sessions	Independent observer
	Attitude	Depression Attitude Questionnaire, 20-item tool	Pre-post intervention	Self-administered
	Knowledge and competency	Case vignettes	Pre-post intervention	Self-administered
	Communication	Calgary-Cambridge, 37-item tool	Pre-post intervention	Psychiatric nurse
	Acceptability, appropriateness and feasibility	2 FGDs and 13 interviews	Immediately post-implementation	Qualitative researchers

Abbreviations: FGD focus group discussion, *MINI* Mini-International Neuropsychiatric Interview

[a] Qualitative interviews supplemented most of the evaluation methods

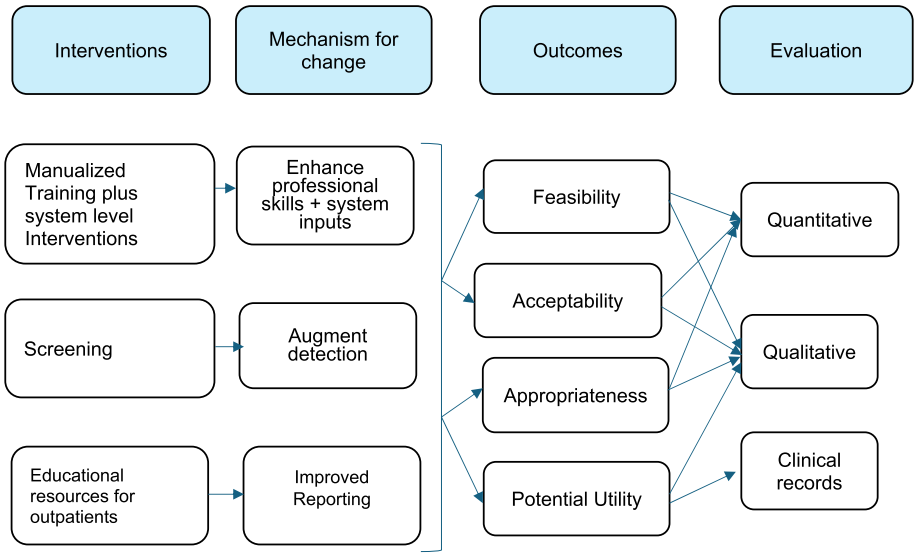

Fig. 3 Intervention package components and process outcome evaluation

assessing primary care clinician. Diagnostic information was extracted daily by project staff from outpatient clinical notes. Confirmatory diagnosis was carried out by psychiatrists, guided by the Mini-International Neuropsychiatric Interview (MINI) [30], within 2 weeks of diagnosis by PHC clinicians.

Implementation process outcomes

Change in knowledge, competency and attitudes Ten case vignettes were used to assess the knowledge and competency of PHC clinicians focusing on common clinical symptoms and diagnoses of depression, risk assessment, pharmacotherapy, brief supportive engagement and follow-up. Change in attitude was evaluated using the Depression Attitude Questionnaire (DAQ), a 20-item tool that evaluates change in assumptions about the aetiology and treatability of depression [31].

Fidelity of the delivery of the training Fidelity was assessed with observer-rated training fidelity and completion checklist, which was developed by the research team based on the contents of the training manual. The checklist consisted of a comprehensive list of key tasks and activities expected to be delivered through all the 19 sessions of training. Each item was rated as yes, no and partially. Time spent for each activity and task was also recorded. The checklist was completed by a non-trainer staff who sat through all the training.

Clinical communication Patient–clinician communication was assessed using the Calgary-Cambridge guide, a 37-item clinical encounter form that measures how the clinician initiates clinical engagement, gathers information and closes the clinical encounter [32]. The guide was completed by a trained psychiatric nurse on 30 clinical encounters before training and on 11 after the training.

Acceptability, appropriateness and feasibility Were assessed both quantitatively and qualitatively (Table 1). The quantitative measures explored the opinion of participants about acceptability, appropriateness and feasibility of the intervention components using a 5-point ordinal scale that captured responses that ranged from complete agreement to complete disagreement [33]. Two focus group discussions (FGD), the first with triage nurses lasting about an hour and the second with clinicians and health administrator lasting 3 h, were conducted.

The acceptability of the information leaflets was assessed by interviewing project staff who distributed the leaflets and outpatients who received the leaflets. All interviews were conducted by experienced interviewers (with more than 10 years' experience in mental health research including qualitative studies and with a minimum of Masters qualification) using topic guides. Interviews were audio-recorded with the consent of participants.

We have conducted two quality improvements (QI) with the PHC providers. The purpose of the QI meetings was to assess progress of implementation and to discuss and address any emerging problems. All the providers

who were involved in the intervention provision or implementation attended the meetings.

Sample size and data management

As this was a pilot study, we did not carry out a formal sample size estimation. However, the number of outpatients who participated ($n=1659$) enables the detection of a 5% difference in the recognition of depression between standard care (control arm) and the intervention arms, as well as between the intervention arms at 80% power and 95% confidence with an intra-cluster correlation coefficient (ICC) of 0.12. This ICC was derived from the same population.

Data on basic characteristics of outpatients were extracted daily from clinical records. For PHC clinicians, basic demographic and educational background (sex, professional qualification and previous mental health training) were collected through interviews. Deidentified data were entered electronically using the Open Data Kit software (https://getodk.org/) and exported to Stata (StataCorp; Revision 24 Jul 2019) for analysis.

The main analysis used multilevel logistic regression. Accounting for the variability between clusters (ICC = 0.12), a two-level hierarchical logistic regression model with robust standard errors was used to model the association between the outcome variable, detection of depression, and selected predictors. Robust standard errors were clustered at the second level in the multilevel model. Because of collinearity, some variables, such as the weekly number of cases, sex of the provider and previous mental health training of providers, were dropped from the model.

All FGDs and interviews were transcribed, translated and then imported into OpenCode 4.0 (https://opencode.software.informer.com/4.0/) for thematic analysis. First, RB read and re-read the transcripts and generated codes. AF then reviewed the codes. Themes were then developed by grouping similar codes. The findings were summarised and are presented with illustrative quotes from participants. On rare occasions, we have used Amharic words when we did not feel that the English word fully captured the sentiment of the Amharic word. This may give readers the opportunity to explore the meaning of those words or phrases if they would like to.

Results

Characteristics of outpatients

Over the 4 weeks of the trial period, 2458 outpatients, with 1659 outpatients fulfilling inclusion criteria, attended the study facilities (Fig. 4).

Characteristics of providers

Fourteen clinicians (six women and eight men) provided the interventions: two in Level-I, four in Level-II, five in Level-III and three in the control facility. Six clinicians were health officers with four non-specialist doctors and four clinical nurse practitioners (one with BSc degree and three diploma level). All clinical staff working in the outpatient clinics and providing clinical assessment and treatment were included as providers irrespective of their qualification. Seven triage nurses conducted the screening (Table 2).

All clinicians participated in the evaluation of attitude, competency and clinical communication. The triage nurses participated in the evaluation of the acceptability and feasibility of screening (Additional file 1: Table S3).

Detection of depression

Of the 1659 outpatients included, 7.0% ($n=116$) had received a diagnosis of depression: 20 from Level-I ($n=20/388$; 5.2%), 41 from Level-II ($n=41/407$; 10.1%), 54 from Level-III ($n=54/585$; 9.2%) and one from the control facility ($n=1/279$). Ninety-three participants (80.2%), who received a PHC diagnosis of depression were available for confirmatory diagnosis by psychiatrists: 19/20 (95.0%) from Level-I, 34/41 (82.9%) from Level-II and 40/54 (74.1%) from Level-III facilities. Overall concordance between PHC and psychiatrist diagnosis was 82.8% ($n=77/93$) and comparable across the intervention facilities.

Detection of depression was associated with receiving one of the three interventions (Table 3). In the fully adjusted model, compared to the control facility, the odds of receiving a diagnosis of depression were over tenfold for the Level-I intervention facility (AOR = 11.75; 95% CI = 5.98, 23.11), rising to over 20-fold for Level-II (AOR = 26.90; 95% CI = 14.63, 49. 47) and Level-III facilities (AOR = 24.95; 95% CI = 13.98, 44.54). Level-II and III facilities performed significantly better than the Level-I facility, without significant difference between the two.

Sixty-two out of 116 people who were diagnosed with depression were started on antidepressant medications: one at the control site, 14 (70%) at Level-I, 25 (61%) at Level-II and 22 (40.7%) at Level-III.

Screening and detection

About half of the outpatients who were administered the screening questionnaire had screened positive. However, on average, only about one in five screen positives were diagnosed with depression by PHC clinicians (Additional file 1: Tables S4–S5). As individual items, irritability and noise intolerance have good sensitivity and negative predictive value, comparable with depressed

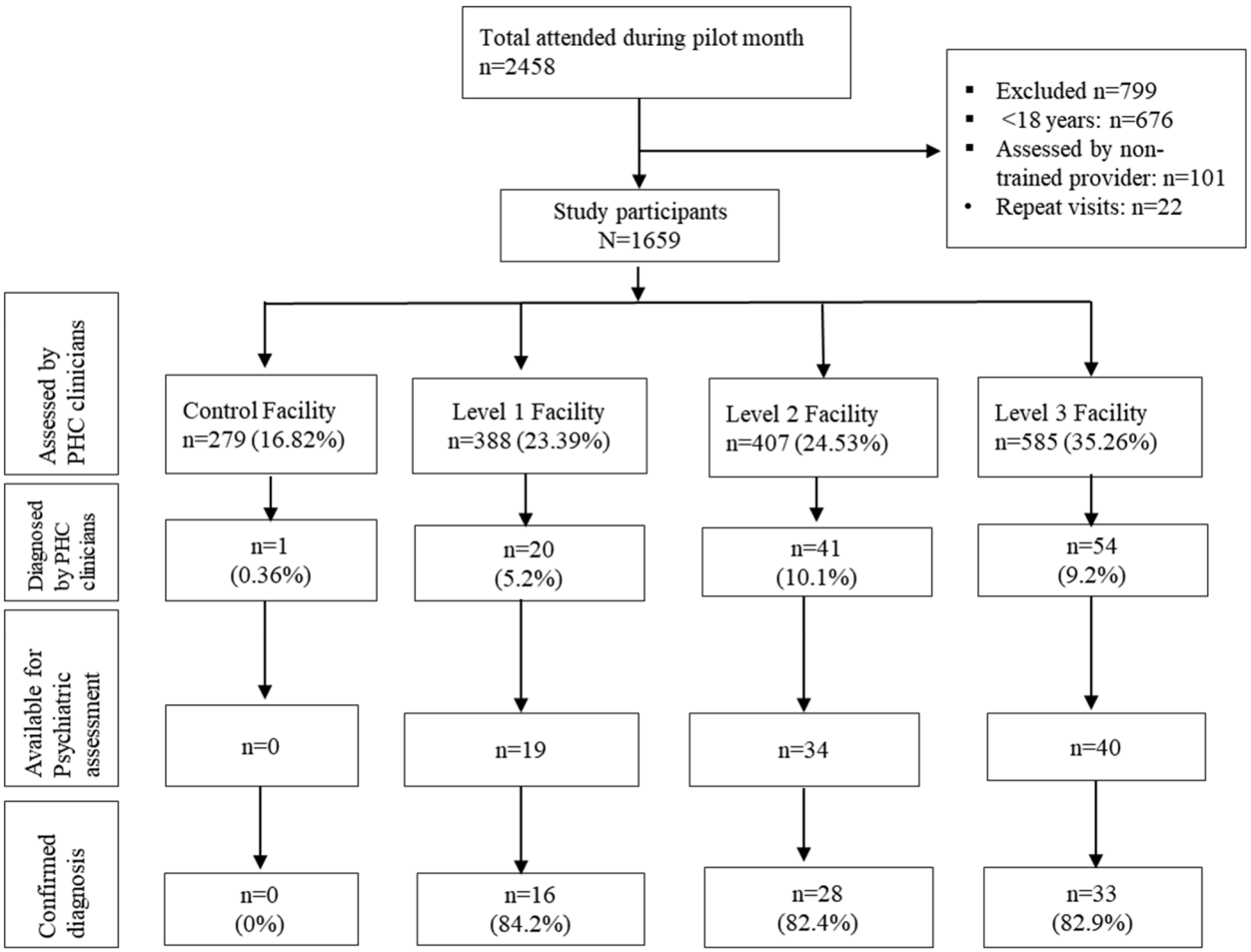

Fig. 4 Participant flow diagram. *Detection is reported based on diagnosis by PHC clinicians, not based on the confirmatory diagnosis

mood (Additional file 1: Table S4). The original items of depressed mood and loss of interest have marginally better specificity and positive predictive value. None of the items or their combination has optimum performance in terms of positive predictive value. Loss of interest was considered the most difficult to enquire about by the clinicians and had low sensitivity compared to the other items.

Implementation process outcomes

Acceptability of the training package and fidelity of delivery

All 5 days of the lecture-based sessions received a very good rating for acceptability (Additional file 1: Table S6). The participatory nature of the training was appreciated, and trainees felt that the training was easy to understand. The length of training sessions was mostly rated as very good or adequate—few thought the training on days 2 and 3 was too long.

The fidelity checklist confirmed that all activities across all sessions during the 5 days of training were completed fully except for the demonstration of the use of the mobile application for 'other' mental disorders, which was replaced by a revision session.

Pre- and post-implementation evaluation of change in knowledge, attitude, competence and clinical communication

There was a significant improvement in the knowledge, attitude, competency and clinical communication skills of providers (Table 4). The greatest improvement was observed in clinical communication skills, followed by improvements in competency and attitudes.

Post-implementation evaluation of feasibility, acceptability and appropriateness of intervention components by providers

The training package good ratings across the four acceptability, feasibility and appropriateness domains of the AIM (Additional file 1: Table S7). There was also good agreement on the acceptability, feasibility and

Table 2 Sociodemographic characteristics of study participants-outpatients and providers

Characteristics		Total *n* (%)	Control *n* (%)	Level I *n* (%)	Level II *n* (%)	Level III *n* (%)	χ^2 (*p* value)
Patient characteristics (*n* = 1659)							
Gender	Male	740 (44.6)	119 (57.4)	190 (51.0)	180 (55.8)	251 (57.1)	4.13 (0.25)
	Female	919 (55.4)	160 (42.6)	198 (49.0)	227 (44.2)	334 (42.9)	
Age (years)	18–24	366 (22.1)	52 (18.6)	75 (19.3)	78 (19.2)	161 (27.5)	68.63 (< 0.001)
	25–34	410 (24.7)	49 (17.6)	101 (26.0)	93 (22.8)	167 (28.5)	
	35–49	473 (28.5)	79 (28.3)	124 (32.0)	104 (25.6)	166 (28.4)	
	50 +	410 (24.7)	99 (35.5)	88 (22.7)	132 (32.4)	91 (15.6)	
	Mean age (SD) years	38.0 (15.9)	41.9 (16.76)	38.3 (15.49)	40.7 (17.17)	34.3 (13.94)	
Residence	Rural	1243 (75.0)	155 (55.6)	388 (100.0)	347 (85.5)	353 (60.3)	276.13 (< 0.001)
	Urban	415 (25.0)	124 (44.4)	0 (0)	59 (14.5)	232 (39.7)	
Health providers' characteristics (*n* = 14)							
Gender	Male	8 (57.1)	2 (66.7)	0 (0.0)	2 (50.0)	4 (80.0)	3.93 (< 0.001)
	Female	6 (42.9)	1 (33.3)	2 (100.0)	2 (50.0)	1 (20.0)	
Educational background	Diploma	3 (21.4)	1 (33.3)	1 (50.0)	1 (25.00)	0 (0.0)	2.62 (0.45)
	Degree	11 (78.6)	2 (66.7)	1 (50.0)	3 (75.00)	5 (100.0)	
Qualification	General practitioner	4 (28.6)	0 (0.00)	0 (0.0)	1 (25.00)	3 (60.0)	9.53 (0.15)
	Health officer	6 (42.9)	2 (66.7)	0 (0.0)	2 (50.00)	2 (40.0)	
	Nurse	4 (28.6)	1 (33.3)	2 (100.0)	1 (25.00)	0 (0.0)	
Year of service	Mean (SD)	6.21 (6.49)	10.33 (4.51)	1.5 (0.7)	8.7 (10.3)	3.6 (2.88)	
Previous mental health training (mhGAP-IG)	Yes	4 (28.6)	3 (100.0)	0 (0.0)	2 (50.0)	0 (0.0)	5.833 (0.12)
	No	10 (71.4)	0 (0.0)	2 (100.0)	2 (50.0)	5 (100.0)	
Health system							
Average number of patients per week		615	91	128	166	230	91.618 (< 0.001)

Abbreviations: mhGAP mental health Gap Action Programme intervention guide

Table 3 Factors associated with detection of depression: multilevel logistic regression

Factors		Crude odds ratio (COR) (95% CI)	*p* value	Adjusted odds ratio (AOR)* (95% CI)	*p* value
Intervention	Control	Ref	Ref	Ref	Ref
	Level-I**	15.23 (10.02, 23.15)	< 0.001	11.75 (5.98, 23.11)	< 0.001
	Level-II**	31.98 (20.01, 51.08)	< 0.001	26.90 (14.63, 49.47)	< 0.001
	Level-III**	28.88 (19.31, 43.19)	< 0.001	24.95 (13.98, 44.54)	< 0.001
Patient-related factors (*n* = 1659)					
Sex	Male	Ref	Ref	Ref	Ref
	Female	1.99 (1.40, 2.82)	< 0.001	2.00 (1.41, 2.83)	< 0.001
Age (years)	18–24	Ref	Ref	Ref	Ref
	25–34	0.87 (0.56, 1.69)	0.65	0.83 (0.54, 1.27)	0.39
	35–49	1.28 (0.87, 1.89)	0.33	1.28 (0.90, 1.82)	0.17
	50 +	0.97 (0.57, 1.67)	0.90	0.99 (0.59, 1.69)	0.98
Residence	Urban	Ref	Ref	Ref	Ref
	Rural	1.37 (0.89, 2.09)	0.15	1.39 (0.97, 2.01)	0.08
Health provider-related factors (*n* = 14)					
Year of service	> 5 years	Ref	Ref	Ref	Ref
	< 5 years	1.81 (0.48, 6.72)	0.38	0.85 (0.54, 1.33)	0.47
Qualification	General practitioner	Ref			
	Health officer	0.51 (0.17, 1.47)	0.21		
	Nurse	0.73 (0.32, 1.67)	0.45		

Intra-cluster correlation coefficient for the full model = 0.0024

*Adjusted for patients' sex, age, residence and providers' years of service. Qualification was omitted from the full model because of a high degree of confounding. The sex of providers was omitted from the model because of perfect collinearity with the qualification of providers (tetrachoric rho = 1.0)

**Difference between the 3 active intervention groups in the fully adjusted model: Level-I and Level-II significantly different (*p* value < 0.0001); Level-I and Level-III significantly different (*p* value < 0.0001); no significant difference between Level-II and Level-III (*p* value = 0.870)

Table 4 Change in knowledge, competency, attitude and communication following training

	Pre-training		Post training		Mean difference (post–pre training)		*p* value (*T*-test)
	Mean	95% CI	Mean	95% CI	Mean	95% CI	
Knowledge	5.97	5.90, 6.03	6.16	6.10, 6.23	0.19	0.12, 0.27	< 0.01
Competence	5.24	5.15, 5.33	9.35	9.26, 9.44	4.11	4.00, 4.21	< 0.01
Attitude	48.19	47.82, 48.56	53.01	52.74, 53.28	4.82	4.59, 5.05	< 0.01
Communication	18.70	18.54, 18.85	34.10	33.79, 34.41	15.4	15.12. 15.68	< 0.01

Abbreviations: 95% CI 95% confidence interval

appropriateness of the screening amongst triage nurses implementing the screening (Table 5).

Overall acceptability of intervention and integrated care for depression

Both the overall acceptability of the programme or implementation of the integrated care for depression as well as the acceptability of the individual components of the intervention packages were assessed. In terms of overall acceptability, all participants in the qualitative study expressed a high level of satisfaction with the main intervention components (training package, decision support and screening). The intervention was considered relevant for the broader community. One of the participants of the clinician FGD stated '*In truth, what has been done in the past four weeks was big. We understood we had a problem with the detection of depression. We understood we were not there for our population.*' [ID 15].

Another participant said '*We have been able to see the problem within our community. We had ignored the problem for far too long. We just ordered laboratory tests and did things that were not appropriate for the disease. We focused on communicable diseases*; [...] *we*

Table 5 Acceptability, feasibility and appropriateness of screening (triage nurses and clinicians)

	Number (percent)					Mean
	Completely disagree	Disagree	Neither agree nor disagree	Agree	Completely agree	
Acceptability of Intervention Measure (AIM)						
Screening meets my approval	0 (0)	1 (6)	0 (0)	4 (25)	11 (69)	4.74
Screening is appealing to me	0 (0)	0 (0)	0 (0)	5 (31)	11 (69)	4.81
I like screening	0 (0)	0 (0)	0 (0)	4 (25)	12 (75)	4.85
I welcome screening	0 (0)	0 (0)	0 (0)	3 (19)	13 (81)	4.89
Feasibility of Intervention Measure (FIM)						
Screening seems implementable	0 (0)	0 (0)	0 (0)	10 (63)	6 (38)	4.56
Screening seems possible	0 (0)	0 (0)	0 (0)	10 (63)	6 (38)	4.59
Screening seems doable	0 (0)	0 (0)	0 (0)	9 (56)	7 (44)	4.67
Screening seems easy to use	0 (0)	0 (0)	0 (0)	8 (50)	8 (50)	4.89
Intervention Appropriateness Measure (IAM)						
Screening seems fitting	0 (0)	0 (0)	0 (0)	3 (19)	13 (81)	4.89
Screening seems suitable	0 (0)	1 (6)	0 (0)	2 (13)	13 (81)	4.78
Screening seems applicable	0 (0)	0 (0)	0 (0)	9 (56)	7 (44)	4.59
Screening seems like a good match	0 (0)	0 (0)	1 (6)	5 (31)	10 (63)	4.70

thought it [depression] *was a problem of White people and not a problem in our country.*' [ID 12].

Another clinician added '*We* […] *have found a vaccine against depression…because it [the intervention] has allowed us to find what was hidden in the past.*' [ID 13].

They also affirmed that it was even preferable to provide care for depression integrated within PHC. '*Treating depression like any other condition has more advantages. If a person with depression is referred to a specialist, [*the person*] will be more uncomfortable; will have concern about confidentiality. There will be mistrust and refusal*' [ID 11]. Other clinicians mentioned that the intervention has led to '*improvement in the quality of service*' [ID 12] and enabled '*us to work with better quality*' [ID 15].

We have seen this clearly […] *without this study, no one would come and report about their mental health problems. In fact, we were talking about it the other day. Many would come complaining of diarrhoea, headache, or such other thing… Without knowing they have depression they would get into more problems. Through this study, it was possible to identify many with depression* [ID 01].

The perceived utility and challenges of implementation were also highlighted in the quality improvement meetings (Additional file 1: Fig. S2).

Training package

The training package was described by the clinicians who participated in the FGD as '*short, precise and applicable and [*that*] the core points were surprisingly well [*presented*]*' [ID 08]. '*We understood how good the training was when we got to our place of work. In the past we detected cases only when* [the illness] *was severe or when the patient was suicidal*' [ID 10]. Clinicians appreciated the role of service users in the training, whose contribution extended '*from assessment and communication skills to initiation of treatment and provision of psychoeducation*' [ID 11]. '*[…] It helped us to associate what we learnt in theory with practice*' [ID 10]. '*Being involved in the practical training in addition to the theoretical training has helped us to complete the course well equipped*' [ID 13].

System interventions (mobile application and posters)

There was a unanimous view that the mobile application (https://abugida.netlify.app/) was very useful in terms of assisting practice. Participants noted the accessibility of phones and the design of the app to be important features. '*We carry our phones all the time. Therefore, the mobile app was more helpful than the training manual*' [ID 10]. They found the structure and navigability also friendly. '*We follow the symptoms through impairment. We also used it [the App] in some patients with psychosis. Once we memorised this, in the third and fourth week, the application is less important*' [ID 06]. While the utility of the app was emphasised, the clinicians also made some recommendations for enhancing the utility of the app, for example, developing an *Amharic* (local language) version of the app as the language of interview. '*Sometimes the words are not easy to translate immediately (in the clinical encounter context)*' [ID10]. The users did not come across any complaints or concerns expressed by outpatients about the use of the app.

Participants considered the posters to be useful in providing quick and easy reference and were also less distracting. Clinicians also noted that the posters reduced dependence on trained staff as these were posted and available in each outpatient clinic. In this regard '*the poster is probably more useful than we think*' [ID 13].

Screening

The perspective of triage nurses Triage nurses were convinced that screening had increased recognition of depression and reduced the workload of clinicians. It was also acceptable to outpatients.

One of the triage nurses participating in the post-intervention focus group said '*The screening was great (ኪሪፋ). We were able to identify many cases. The questions were clear and simple to administer*' [ID 07]. Others noted '*… Our community does not come saying "I have this* [depression] *problem". Using the screening, we were able to identify many cases.*' [ID 03].

'*Sometimes, when we ask them, the patients even say,* "*How did you know…who told you that I have these problems*?" *I am very happy*' [ID 05]. They also reported that the screening questions were easy to understand and ask. '*The questionnaire was also very clear*' [ID 07].

Perspective of clinician providers Most providers considered the screening as '*critical/crucial* (ወሳኝ ነው)' [ID 13]. The interest of the clinicians was primarily in the reduction of workload by allowing the '*[…] clinicians to focus only on the positive cases*' [ID 08]. Others also mentioned how the screening may have enhanced community awareness about the service: '*People knew about the service because of the screening*' [ID 09]. One clinician gave an example of a patient who brought his brother from Addis Ababa, the capital city 100 km away, to be evaluated because the patient had the experience of being screened in the first place.

Information leaflets

There was clear interest in the information leaflets by outpatient attendees. No participant refused to accept the leaflets. Some attendees recommended distribution of the leaflets to outpatients with long-term conditions and those attending maternal health services. A similar opinion was shared by clinician providers, who even recommended for the content of the leaflets to be broadcast on national TV. They reported seeing people reading the leaflets or having them read by another. One participant who had some depressive symptoms claimed feeling better after reading the leaflet. '*The leaflet itself counsels and heals (ራሱ መካሪ ነው። ራስን ያድሳል።)*' [Service User IDI 06]. Other participants agreed, for example, Service User IDI 05 mentioned '*በጣም በጣም ጠቀሜታ አለው*/*betam betam tekemeta alew*', i.e. (*The leaflet* is *very very* useful). Another suggested that the '*leaflets should be distributed to the community…*' [Service User IDI 01].

Acceptability of diagnosis for outpatients

Outpatients were generally happy to be asked about depressive symptoms. One clinician asserted that '*Patients were all happy* [to be asked about depression]*-both those with and without depression. Even some brought other patients*' [ID 09]. Another commented, '*…I have not come across anyone who complained about the diagnosis. In fact, after psychoeducation, they understand it is something they can link up with their problems…there was positive expectancy in all of them.*' [ID 15].

However, two clinicians reported encountering outpatients with concerns about the diagnosis. The first clinician [ID 08] reported of two outpatients who were '*shocked*' by the diagnosis primarily because they did not think depression would get better and could even lead to rejection by their family. The second clinician [ID 01] reported a more severe response. '*It was not pure depression…he* [patient] *knew M* [the psychiatry nurse working at the facility]. *When I asked for help from M*, [the patient said] *"I am not M's case…I have a headache, I know it, I always have typhoid fever and typhus. I came to take medications and be treated* [for these problems]*". He finally refused to be treated and left…he couldn't believe it that he was diagnosed with depression.*'

Challenges to implementation

The key challenges or feasibility concerns raised in the implementation were the workload, weak coordination and insufficient inclusivity.

The perceived complexity and time-consuming nature of assessments was of concern to clinicians.

'*What takes time is the assessment. We need to dig up…it could be another medical problem…malnutrition, anaemia, hypertension, cardiac problem. We have to do a physical examination to check for these. This takes time. Other patients are waiting for their turn to be seen* [while the assessment goes on]. *We need to explain. What does loss of interest in pleasurable activities mean? We have to find out what pleasurable activities they used to do. Then we have to ask them if they have stopped doing those things. We also do not want to interrupt them. Patients may also not come on time for their follow-up*' [ID 12].

'*Building trust takes time. Learning about stressors takes time.*' [ID 08].

Diagnosing depression in those with another medical condition was described as '*confusing*'. One clinician [ID 09] mentioned her difficulty diagnosing depression in a patient with brain cancer. Although she felt the person had depression, she was unsure and consulted a psychiatrist, who prescribed antidepressants. '*I believe system change is needed. We must ask how these challenges should be addressed. But the work needs to continue*'.

Another area of concern was the workload from follow-ups: '*We have prepared many for follow-up; probably cannot be covered by just one clinician. For example, I have identified 20 or 21 cases. Except for three (one has changed address and the others were not contactable), the rest have come. I had to see all these*' [ID 08].

For triage staff, administering the screening questionnaire was something they did as part of a broader triage responsibility. This was of '*particular problem in the morning and early afternoon when patient flow was high*' [ID 06]. Clinicians expressed concern that triage nurses would neglect other responsibilities as their workload increased. One clinician noted that '*screening for cough and vital signs was reduced during the IDEAS (Improving Detection of Depression Study) implementation*' [ID 09]. The triage staff also had to coordinate their work with other teams. '*Occasionally, patients did not understand the questions, which required more time to complete the form*' [ID 06]. Training other clinicians, involving more triage staff, including those working in clinical records was recommended. Clinicians also noted that the study was not inclusive enough. For example, patients with chronic conditions in follow-up, who were considered more at risk for depression, and mothers attending maternal and childcare clinics were not part of the study. The clinicians proposed that these services should be included in an integrated care framework.

Discussion

This report is part of a comprehensive consideration of the detection of depression in PHC in a LMIC setting. The interventions tested in this study were co-developed with relevant stakeholders through a bottom-up approach that identified contextually meaningful symptoms and concepts of depression [19, 23, 24, 27–29, 34].

The study has demonstrated that the detection of depression in PHC can be improved substantially through structured multicomponent interventions that target care providers, service users and the health system. The intervention appears to be effective irrespective of setting (rural vs. urban) or qualification of providers (nurses, health officers or doctors). Virtually all the elements of the intervention (training, screening, decision support and user educational resources) were considered feasible, acceptable and appropriate. The comparison of the intervention to vaccination made by one provider is somewhat apt and indicative of the perceived utility and acceptability of the intervention. While the practitioners were realistic about the potential increase in workload, they considered the intervention to be an opportunity to provide care for their community rather than as a 'dumping' of mental health care onto their workload [35]. Perhaps of note has been the recognition of somatic phenomenon in depression by the clinicians. It has been long recognised that about two third of people with depression in PHC present with somatic complaints making diagnosis of depression more difficult [2]. To some extent, the qualitative evidence suggests that the clinicians have begun to acquire relatively 'sophisticated' skills to differentiate depression from somatic conditions.

Factors associated with improved detection

The IDEAS intervention was the most important factor that was associated with better detection. Compared to standard care, all the three intervention arms were more effective. Detection in the active intervention facilities was also superior to what has been reported previously within Ethiopia [19] and elsewhere in LMICs [34].

Components of intervention

Training package: Although requiring significant modifications in focus and delivery, the mhGAP-IG contains the key elements for equipping PHC clinicians with the necessary know-how for diagnosing depression. The practical nature of the training and the involvement of service users as trainers were important elements that are in principle replicable in other similar low-income country settings.

Screening: Screening integrated within the triage system appears feasible and acceptable with significant utility. The two facilities where screening was implemented, Level-II and Level-III facilities, had double the level of detection to that of the Level-I facility. The two new items (noise intolerance and irritability) added to the PHQ-2 [36] also appear to have potential utility for use in future studies. With regard to the individual screening items, loss of interest had the lowest performance. It also has lower acceptability by clinicians because of the complexity of asking about it. On the other hand, the new items, noise intolerance and irritability were often volunteered by patients and were easy to enquire about. Overall, the main problem with the screening was the high false positivity rate, with only 20% of screen positives confirmed to have depression. While the promise of screening is substantial, further adaptation of tools is needed to improve performance. Given its complexity and low sensitivity, it may be necessary to drop 'loss of interest' from the screening questionnaire. However, it is worth noting that

screening may help to normalise discussion of emotional problems.

System level interventions: Both the mobile application and posters were considered useful. The almost universal availability of mobile phones and flexibility of the application were important elements. The mobile application also takes advantage of the increasing technology dependence of the clinicians. Although the app was incorporated because of the suggestion of PHC clinicians, the recommendation for developing an Amharic (local language) version of the app was made after the need for such a version was identified after the practical application of the app in the English version. Next iterations should develop the Amharic version. The poster was perhaps an 'undervalued' tool, as the clinicians suggested. These tools were developed from the training materials of the WHO mhGAP-IG, creating synergy with the mhGAP-IG resources. Both tools are easy to replicate in other LMIC settings.

User awareness raising: Although no additional benefit was observed in the Level-III facility, there was a general perception from service users and providers that the leaflets were useful. Further evaluation of this component is warranted.

The major limitation of the study was the small number of clusters. While this may be acceptable for a pilot study, and that the issue of power did not arise for the primary outcome of detection of depression, it has affected our capacity to explore differences by various factors. For example, the study facilities were relatively diverse with variation in the qualifications of practitioners and rural-urbanicity. Although this reflects the reality on the ground, exploring the impact of these factors on detection is important for optimising the intervention. Thus, the work must be replicated in a larger number of randomised clusters. The non-randomised allocation is another important consideration with potential risk of introducing bias and confounding. However, the size of the difference and consistency across the active intervention arms suggests that the intervention packages were important in explaining the higher rate of detection in the intervention arms. Nevertheless, a larger-scale randomised trial is indicated. While the risk of contamination is an important consideration, the health facilities were far apart. The PHC clinicians have also highlighted the occasional challenge of using a mobile application developed in English. This may lead to inconsistent application of the words in the app. However, we do not think this would exaggerate the level of detection. Subsequent iterations of the app should take this into account. Cost implications of the interventions also need to be determined. As much as we have attempted to understand the contribution of the individual components of the intervention, there could have been duplication and redundancy. Furthermore, the role of addressing user level barriers was not explored sufficiently.

Conclusions

In the context of the very limited number of studies that improved detection of depression in PHC, this study may have brought an important assurance that effective scale up of mental healthcare for people in LMICs is possible. The WHO mhGAP-IG is an important but not a sufficient tool for improving detection of depression—additional components are required. Although this study has demonstrated that PHC clinicians can recognise depression effectively, the attendant work burden should be anticipated, and the right preparations made. Screening with a culturally relevant tool has promise; however, further optimisation is required. Detection also does not necessarily mean that people with depression receive the right care. Further focused work is required not only to improve detection, but also to equip clinicians with the skills for ongoing care with additional system supports, such as supervision and referral pathways. A pragmatic large-scale randomised study is needed to support the cost-effective scale up of high-quality integrated care for people with depression.

Abbreviations

AIM	Acceptability of Intervention Measure
DAQ	Depression Attitude Questionnaire
FGD	Focus Group Discussion
FIM	Feasibility of Intervention Measure
IAM	Intervention Appropriateness Measure
ICC	Intra-cluster Correlation Coefficient
IDEAS	Improving Detection of Depression Study
LMICs	Low- and Middle-Income Countries
mhGAP-IG	Mental Health Gap Intevention Guide
MINI	Mini-International Neuropsychiatric Interview
MRC	Medical Research Council
PHC	Primary Healthcare
PHQ	Patient Health Questionnaire
QI	Quality Improvement
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04010-4.

Additional file 1: Fig. S1, Tables S1–S7, Fig. S2. Fig. S1 Process of intervention development and testing. Table S1 Process of intervention evaluation using Bowen's Feasibility Framework. Table S2 Health facilities in the study district and population served. Table S3 Participants in the qualitative study on feasibility and acceptability. Table S4 Performance of practitioners and screening by qualification and intervention level. Table S5 Sensitivity and specificity of screening items against clinician diagnosis. Table S6 Training satisfaction (clinicians). Table S7 Acceptability, feasibility and appropriateness of training (clinicians). Fig. S2 Summary of feedback from quality improvement sessions.

Additional file 2: Consort extension for pilot and feasibility trials checklist.

Acknowledgements
We acknowledge the contributions of all stakeholders who delivered the intervention and participated in the pilot trial. We thank the PHC providers for sharing their time and experiences during the study. CH receives support from the National Institute for Health and Care Research (NIHR) through the NIHR Global Health Research Group on Homelessness and Mental Health in Africa (NIHR134325) and the SPARK project (NIHR200842), using UK aid from the UK Government. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. CH also receives support from the Wellcome Trust through grants 222154/Z20/Z and 223615/Z/21/Z.

Authors' Twitter handles
Twitter handles: @Abebawfekadu (Abebaw Fekadu),@MartinKGHI (Martin J. Prince), @MaystonRosie (Rosie Mayston), @richinabirhane (Rahel Birhane).

Authors' contributions
AF and MJP conceptualized the study and AF led the study. GM, CH, HL, AT, AA, BT, RM, AJC, IP, GT, CL, RM and MD provided specific methodological expertise. AF, MJP, CH, BM, RB, KH and GM designed the intervention content and data collection instruments. AF, RB, TG, TE, BM and MY contributed to intervention delivery and data collection. AF, RB, MB, EA and GM had full access to the data. AF and RB accessed and verified the data. AF, RB, EA, GM and MB did the statistical analyses. AF wrote the first draft of the report, and all authors reviewed and approved the final version.

Funding
This study was financially supported by UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordant agreement through the Africa Research Leader scheme (Grant number MR/M025470/1). The funders had no role in the study design, data collection and analysis, decision to publish or manuscript preparation.

Data availability
De-identified participant data will be made available after publication, up on a reasonable request to the corresponding author, AF.

Declarations

Ethics approval and consent to participate
The authors declare that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human patients were approved by the Addis Ababa University Institutional Review Board (reference number 007/18/Psy (2018–2021)). Written informed consent was obtained from each participant.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Center for Innovative Drug Development and Therapeutic Trials for Africa (CDT-Africa), Addis Ababa University, Addis Ababa, Ethiopia. [2]WHO Collaborating Center for Mental Health Research and Capacity Building, Department of Psychiatry, School of Medicine, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia. [3]Department of Global Health and Infection, Brighton and Sussex Medical School, Brighton, UK. [4]Health Service and Population Research Department, Institute of Psychiatry, Psychology, and Neuroscience, Centre for Global Mental Health, King's College London, London, UK. [5]Department of Inflammation Biology, Centre for Rheumatic Diseases, Faculty of Life Sciences and Medicine, King's College London, Cutcombe Road, 10, Cutcombe Rd, Weston Education Centre, London SE5 9RJ, UK. [6]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia. [7]School of Psychology, College of Education and Behavioral Studies, Addis Ababa University, Addis Ababa, Ethiopia. [8]Centre for Rural Health, School of Nursing and Public Health, University of Kwazulu-Natal, Howard College, Mazisi Kunene Road, Durban 4001, South Africa. [9]Centre for Affective Disorders, Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [10]Health Economics and Policy Research Unit, Wolfson Institute of Population Health, Queen Mary University of London, London, UK. [11]Department of Psychiatry, College of Health and Medical Sciences, Haramaya University, Dire Dawa, Ethiopia. [12]Division of Psychiatry, Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK.

Received: 10 September 2024 Accepted: 13 March 2025
Published online: 28 March 2025

References

1. Bromet E, Andrade L, Hwang I, Sampson N, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011;9(1):90.
2. Tylee A, Gandhi P. The importance of somatic symptoms in depression in primary care. Prim Care Companion J Clin Psychiatry. 2005;7:167–76.
3. Ustun TB, Sartorius N. Mental illness in general health care. Chichester, UK: John Wiley; 1995.
4. GBD 2019 Mental Disorders Collaborators. 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. Lancet Psychiatry. 2022;9(2):137–50.
5. Kohn R, Saxena R, Levav I, Saraceno B. The treatment gap in mental health care. Bull WHO. 2004;82:858–66.
6. The WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA. 2004;291:2581–90.
7. Perianayagam A, Prina M, Selvamani Y, Gudekar D, Salvi S, Varghese M, et al. Sub-national patterns and correlates of depression among adults aged 45 years and older: findings from wave 1 of the longitudinal ageing study in India. Lancet Psychiatry. 2022;9(8):645–59.
8. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. BJPsych OPen. 2017;210(2):119–24.
9. Jenkins R, Kiima D, Okonji M, Njenga F, Kingora J, Lock S. Integration of mental health into primary care and community health working in Kenya: context, rationale, coverage and sustainability. Ment Health Fam Med. 2010;7(1):37–47.
10. Tadesse GL, Giorgis TW, Gerba H. Policies, delivery models, and lessons learned from integrating mental health and substance abuse services into primary health care in Ethiopia. FASEB bioAdv. 2021;3(9):694–701.
11. Hanlon C, Medhin G, Dewey ME, Prince M, Assefa E, Shibre T, et al. Efficacy and cost-effectiveness of task-shared care for people with severe mental disorders in Ethiopia (TaSCS): a single-blind, randomised, controlled, phase 3 non-inferiority trial. Lancet Psychiatry. 2022;9(1):59–71.
12. Scott J, Thorne A, Horn P. Effects of a multifaceted approach to detecting and managing depression in primary care. BMJ. 2002;325:951–4.
13. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. Br J Gen Pract. 2011;61(583):e72-80.
14. Alexander CL, Arnkoff DB, Kaburu AW, Glass CR. Detecting depression in rural primary care clinics in central Kenya: impact of a brief training intervention. Int Perspect Psychol. 2013;2:14–28.
15. Udedi M. The prevalence of depression among patients and its detection by primary health care workers at Matawale Health Centre (Zomba). Malawi Med J. 2014;26:34–7.
16. Jenkins R, Othieno C, Okeyo S, Kaseje D, Aruwa J, Oyugi H, et al. Short structured general mental health in service training programme in Kenya improves patient health and social outcomes but not detection of mental health problems - a pragmatic cluster randomised controlled trial. Int J Ment Health Syst. 2013;7(1): 25.
17. Jenkins R, Kiima D, Njenga F, Okonji M, Kingora J, Kathuku D, et al. Integration of mental health into primary care in Kenya. World Psychiatry. 2010;9(2):118–20.
18. Kauye F, Jenkins R, Rahman A. Training primary health care workers in mental health and its impact on diagnoses of common mental disorders

in primary care of a developing country, Malawi: a cluster-randomized controlled trial. Psychol Med. 2014;44(3):657–66.
19. Fekadu A, Medhin G, Selamu M, Wolde-Giorgis T, Lund C, Alem A, et al. Recognition of depression by primary care clinicians in rural Ethiopia. BMC Fam Pract. 2017;18(1):56.
20. Ayano G, Assefa D, Haile K, Chaka A, Haile K, Solomon M, et al. Mental health training for primary health care workers and implication for success of integration of mental health into primary care: evaluation of effect on knowledge, attitude and practices (KAP). Int J Ment Health Syst. 2017;11:63.
21. Gilbody S, Whitty P, Grimshaw J, Thomas R. Educational and organizational interventions to improve the management of depression in primary care: a systematic review. JAMA. 2003;289(23):3145–51.
22. Gilbody SM, Whitty PM, Grimshaw JM, Thomas RE. Improving the detection and management of depression in primary care. Qual Saf Health Care. 2003;12:149–55.
23. Habtamu K, Birhane R, Demissie M, Fekadu A. Interventions to improve the detection of depression in primary healthcare: systematic review. Syst Rev. 2023;12(1):25.
24. Demissie M, Birhane R, Hanlon C, Eshetu T, Medhin G, Minaye A, et al. Developing interventions to improve detection of depression in primary healthcare settings in rural Ethiopia. BJPsych Open. 2024;10(2): e52.
25. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. BMJ : Br Med J. 2015;350:h1258.
26. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med. 2009;36(5):452–7.
27. Mayston R, Frissa S, Tekola B, Hanlon C, Prince M, Fekadu A. Explanatory models of depression in sub-Saharan Africa: synthesis of qualitative evidence. Soc Sci Med. 2020;246: 112760.
28. Tekola B, Mayston R, Eshetu T, Birhane R, Milkias B, Hanlon C, et al. Understandings of depression among community members and primary healthcare attendees in rural Ethiopia: a qualitative study. Transcult Psychiatry. 2023;60(3):412–27.
29. Habtamu K, Birhane R, Medhin G, Hanlon C, Fekadu A. Psychometric properties of screening questionnaires to detect depression in primary healthcare setting in rural Ethiopia. BMC Prim Care. 2022;23(1):138.
30. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59(Suppl 20):22–33 quiz 4–57.
31. Mbatia J, Shah A, Jenkins R. Knowledge, attitudes and practice pertaining to depression among primary health care workers in Tanzania. IJMHS. 2009;3:1–6.
32. Kurtz SM, Silverman JD. The Calgary—Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. Med Educ. 1996;30(2):83–9.
33. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci. 2017;12(1):108.
34. Fekadu A, Demissie M, Birhane R, Medhin G, Bitew T, Hailemariam M, et al. Under detection of depression in primary care settings in low and middle-income countries: a systematic review and meta-analysis. Syst Rev. 2022;11(1):21.
35. Petersen I. Comprehensive integrated primary mental health care for South Africa. Pipedream or possibility Soc Sci Med. 2000;51(3):321–34.
36. Kroenke K, Spitzer RL, Williams JBW. The patient health questionnaire-2: validity of a two-item depression screener. Med Care. 2003;41(11):1284–92.

Publisher's Note