PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol of a randomized clinical trial to integrate mental health services into primary care for post-conflict populations in Northern Sri Lanka (COMGAP-S)
AUTHORS	Doherty, Shannon; Dass, Giselle; Edward, Anne; Stewart, Robert; Roberts, Bayard; Abas, Melanie

VERSION 1 – REVIEW

REVIEWER	Ameel Al Shawi University of Fallujah
REVIEW RETURNED	18-May-2021

GENERAL COMMENTS	Dear authors Thanks for this important protocol that aim to integrate the mental health with primary health care for post conflict population, though there are many points that may require more details and clarification especially in methods such as randomization process, cluster design , date of study conduction …. as following: Abstract It is better to insert some points about ethical issues in the abstract .. Introduction More points for rationale of the study might be required in the introduction section methods and analysis The authors stated in the beginning of methods the aims of the project; it is better to insert such paragraph in the end of the introduction! Primary and secondary outcomes might be required to delete from methods section Details are required for groups in clinical trial. More clarification and details are required for randomization process in order to clarify several important points, as following ? How the author made the randomization ? How they divided the groups ? A diagram may be required for randomization !! Details are required for a stepped wedge cluster design! The authors stated: “Facilities are randomized using standardized randomization techniques to allocate facilities to the training sequence within the stepped wedge design. Patients and community representatives were not randomized. “ such statement needs more explanation and clarification Details are required for the training material that was used for training the primary care practitioners and public health

	professionals ??? I suppose a detail is required for interventions among study groups !! How the authors chose the patients to participate in the study ?? how they obtained the consent from patients to participate in the study ?? more clarification might be needed for such important issue! On which basis the researchers chose three days of training, is it sufficient time for training such important issue ??! Research team : who are the research team, what are the scientific background and experiences of such team ?? Details are required for statistical methods including the variables and the program that would be used for statistics analysis, why they chose generalized linear regression to conduct such analysis? Missing information regarding the date of study conduction. What are the limitations for such project ? References Some references need to rewrite according to the style of the journal.

REVIEWER	Kimberly Hook Boston Medical Center, Psychiatry
REVIEW RETURNED	19-May-2021

GENERAL COMMENTS	Thank you for the opportunity to review this protocol. I believe this is an important area of study and commend the authors for their work. Below, please find my notes. Aims: • Was there a rationale for the percentage targets identified in the aims? For example, why was 50% chosen as a target for stigma reduction? o I think stigma reduction is an interesting outcome and an important one to consider. • The first aim appears to target increasing referrals to "specialist care for mental disorders." My understanding is that the purpose of this intervention is to provide mental health care in primary care settings, not refer to outside providers. Please provide clarity. • The second aim states that "qualitative interviews" will be used to determine concordance between PHC clinicians and psychiatrists. Please provide details regarding what this means. Is this a structured clinical interview of some sort? Study Design: • May also note that using a cluster design can help to minimize contamination at sites. • How were the clinics randomized? Were they matched on any sort of criteria (e.g., size, location)? • The details about timing is well-articulated and clear. Intervention: • Please provide brief details on the content of the training that the various participants received. I am unclear if they were trained on med prescribing, some sort of psychosocial intervention, etc. How was the training delivered – didactics, role plays, etc.? o I appreciate the refresher courses; I think this is important to assist with retaining knowledge. • Please provide brief details on the adaptation process of

	mhGAP. Was it simply translation? Were there any cultural/clinical/structural adaptations needed? • Were the measures adapted or validated in any way? • Please provide additional clarity about the roles of the various professionals engaged in the study. At this time, I am unclear about their functions: o What is the role of public health professional in this study? Are they providing some sort of mental health intervention? What is the training that they receive? o What is the role of the community representative? Are they trying to increase referrals? Is the training mainly to reduce provider stigma? With whom do they engage – patients, providers? • Please provide some details about how any safety issues (e.g., suicidality) will be managed. • What is "supervision" in the context of this intervention? Who provides supervision? Analysis • If possible, including any information about measure validation would be very beneficial. • Were there any fidelity checks/measures that can be included in the analyses? • Additional details about planned analyses are needed (e.g., reaching saturation, statistical approach, etc.). Minor Notes • There are some minor typos/grammatical errors that may be improved with a close review. • Please provide the definition for the AMIQ the first time it is used in the text.

REVIEWER	James Griffin University of Warwick, Warwick Medical School, Clinical Trials Unit
REVIEW RETURNED	13-Jul-2021

GENERAL COMMENTS	The authors have clearly identified an important research question, in a much-needed clinical area. The stepped wedge design could be an appropriate design and it's very welcome to see it being used but it is under described and there are large gaps in the reporting of the research plan. Overall (general) comments: The statistical analysis section needs more detail and would benefit from input by a suitably qualified medical statistician. From reading the full manuscript I can understand what your intended overall aims are. However, a methods paper like this should have much more specific detail regarding the planned analysis – particularly in a clinical trial setting where pre-specification is so important. For example there is no explicit mention in the analysis section as to what variables are included in the analysis? Will inference be drawn from an adjusted or unadjusted model? How will you account for clustering (correlation) of individual patients within facilities? The design of the study is a cluster randomised experiment and as such the primary unit of analysis is the facility. I think the authors should be more careful to highlight this, and be wary of individual level analysis as you may make inappropriate inferences from patient level analysis if the clustering is unaccounted for appropriately. The authors should specify a single primary outcome" rather than listing more than one. I understand the 3 choices reflects the

	different groups of participants that you hope the intervention impacts upon. But there should be a single primary outcome measure linked to the study design and sample size and remaining should be moved to secondary outcomes. The diagrams provided are helpful. They could be improved by being combined in to a single, simpler, schematic so a reader can see the overall picture. There is no mention of key statistical challenges that this trial, like any other trial of a complex intervention, may encounter. These includes section on, but not limited to; missing data, multiple testing, adherence to intervention, planned sensitivity and/or subgroup analyses. Any further revisions of the paper should address these sections. Blinding (masking) is not discussed in the main body of the manuscript. Where blinding is possible it should be detailed. Where it is not possible then a discussion of the impact of this should be presented. Contamination is important to consider in stepped wedge designs – why is there no discussions of potential contamination in the main manuscript? Consider the CONSORT guidelines and relevant extension for stepped wedge designs (https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-3116-3). I would strongly recommend the authors consider this paper as it includes the checklist which the trial will have to report to, and as such the present manuscript should reflect on the areas of the checklist which are not suitably addressed. Minor (specific) comments: Page 4, line 26. The acronym COMGAP has not been defined in full Page 4, line 35, The acronym COMGAP-S has not been defined in full Page 4, line 46, The results presented here are not particularly relevant for the background section and coul be moved to a more appropriate location or better justified. Page 5, line 36. Select a single primary outcome appropriate for the design of a stepped wedge cluster RCT – move the other outcomes to secondary outcomes. Better justification for the selected % changes should be given. Page 6, line 39. What exactly is the unit of analysis/randomisation? The term facilities is not specific enough. Reading the rest of the manuscript a facility could be anything from a district hospital down to a general practice? Specific detail should be given. Page 7, line 5. A list of inclusion/exclusion criteria would be preferable. These be listed separately for facilities and for participants. Page 7, line, 50. Why are only patients with depression and anxiety included? Justification should be given. Page 8, line 47. Refer to the flow diagram to aid explanation. Page 11, line 28. How are facilities enrolled specifically? Page 11, line 30. A table demonstrating the schedule of assessments would be more helpful, following the text is difficult. A table is the recommended methods of displaying these data. Page 12, line 41. This section is incomplete and does not adequately describe the analysis. Page 12, line 52. The sample size is not adequately described in the body of the text. The choice of parameters should be better justified here and explanation given. I suggest describing in detail the process for a single outcome rather than having multiple primary outcomes.

	Page 13, line 33. I would suggest removing this section as it will be out of date and irrelevant by the time of publication. Page 15, There is no discussion of limitations of the current study design and planned analysis.

VERSION 1 – AUTHOR RESPONSE

Response to Reviewer 1

Thank you very much for your comments and suggestions. I apologize if it wasn't clear, but this study was completed as of 30 April 2021 (it was submitted to the journal while the study was being conducted and was completed before reviewer comments were received).

The protocol has gone through a technical review from the funder (Centers for Disease Control and Prevention), and ethics reviews in the UK and Sri Lanka. I have addressed all comments below as best as I could, but am unable to revise some aspects of the protocol as the study has completed. Again, thank you for your careful review and thoughtful comments, it's greatly appreciated.

Reviewer: 1
Dr. Ameel Al Shawi, University of Fallujah
Comments to the Author:

Dear authors
Thanks for this important protocol that aim to integrate the mental health with primary health care for post conflict population, though there are many points that may require more details and clarification especially in methods such as randomization process, cluster design, date of study conduction …. as following:

1. Abstract

It is better to insert some points about ethical issues in the abstract .

The following information about ethics approvals has been added to the abstract (page 2)

Ethics and dissemination

This trial has received ethical approval from the Faculty of Health, Education, Medicine and Social Care, Anglia Ruskin University, UK (SC/jc/FMFREP/16/17 076), from the Faculty of Medicine, University of Jaffna, Sri Lanka (J/ERC/17/81/NDR/0170) and non-engagement approval from the funding body, the Centers for Disease Control and Prevention (2018-015). All participants gave written consent.

Dissemination of study results will be completed through publication of academic articles, conference presentations, townhall meetings, written pamphlets in plain language, reports to Ministry of Health and other government organisations, and through social media outlets.

2. Introduction
More points for rationale of the study might be required in the introduction section

The following additional information has been provided and some information moved to increase clarity of the rationale (page 5)

The first phase of the study was completed in 2015-2016 and facilitated understanding of the mental disorder burden and treatment gap at primary care level. Results from the cross-sectional study indicated the most prevalent mental health disorders in primary care settings were depression (41.1%, 95% CI: 38.7-44.5%), anxiety (46.7%, 95% CI: 41.9-51.5%), post-traumatic stress disorder (13.7%, 95% CI: 10.6-16.8%), and psychosis with hypomania (17.6%, 95 CI: 13.3-21.9) [13]. While this facilitated understanding of the underlying mental health issues within the region, the issue of unmet need was still present. In this context, the aim of Phase 2 of COMGAP-S is to use a scaled-up training intervention based on the World Health Organization Mental Health Gap Action Programme (WHO mhGAP 2.0) to integrate mental health services into primary care by providing training to primary care practitioners and public health professionals serving conflict-affected populations in

primary care settings in the Northern Province [12]. The second phase of the project was conducted between 2016-2021.

3. Methods and analysis
a. The authors stated in the beginning of methods the aims of the project; it is better to insert such paragraph in the end of the introduction!

Thank you, this section has been moved to the end of the Introduction as suggested (pages 6-7)

b. Primary and secondary outcomes might be required to delete from methods section

Thank you, primary and secondary outcomes have been added to the end of the Introduction as suggested (pages 6-7)

c. Details are required for groups in clinical trial.

Additional information has been added for each group included in the clinical trial on the following pages of the manuscript

Pages 11-12
Primary care practitioners (general physicians practicing in selected facilities) will undergo one month of monitoring prior to training to establish a baseline. Primary care practitioners will be trained using the mhGAP manual to recognize common mental health disorders using screening procedures, delivery brief psychosocial interventions (psychoeducation, Motivational Interviewing), and referral procedures for complex cases.
Page 13
Public health professionals (nurse attendants, midwives) in each selected facility will undergo one month of monitoring prior to training to establish a baseline. Public health professionals will be trained using adapted written materials from the mhGAP training to recognize signs of common mental health issues, provide basic psychosocial care (education) and refer onwards if needed.
Page 14
Community representatives (teachers, social workers) will undergo one month of monitoring prior to training to establish a baseline. Community representatives will be trained using adapted materials from the mhGAP training material to recognize signs of common mental illness, delivery basic psychosocial education and learn referral pathways and where to seek providers.

d. More clarification and details are required for randomization process in order to clarify several important points, as following? How the author made the randomization? How they divided the groups? A diagram may be required for randomization!!

Additional information has been added to clarify the randomization process (pages 8-9)

Facilities will be randomized using standardized randomization techniques to allocate facilities to the training sequence within the stepped wedge design. Patients and community representatives will not be randomized. The study team cannot be blinded to the allocation of facilities. However, the study psychiatrist will be masked to the randomization status of facilities from which patients originate. Patients will also be blinded to facility randomization status. [1]

A list of public primary care facilities has been provided to the study team by the Northern Ministry of Health. A random number generator will be used to select 23 facilities across the 5 districts. Distribution of clusters will be allocated proportionally to the total numbers of internally displaced people in each district (districts with larger numbers were assigned more clusters). This will ensure adequate representation of conflict displacement and severity.

e. Details are required for a stepped wedge cluster design!

Additional details have been provided about the stepped wedge cluster design (page 8*)*

This is a randomized clinical trial using a stepped wedge cluster design [14]. This design comprises a preliminary period where no clusters (healthcare facilities) are exposed to the intervention, then at regular intervals, one cluster is randomized to cross from control to intervention. This continues until

all clusters have been exposed [14]. The stepped wedge design is increasingly utilized as a way to evaluate interventions that involve service delivery and involves sequential and random crossover of clusters (facilities) from control (delivery of standard care) to intervention (mhGAP training) until all clusters are exposed [14]. The stepped wedge design takes into account the logistical constraints of needing to sequentially roll out the intervention by training one facility at a time enabling us to understand how this intervention can be implemented in the future on a larger scale. This design was chosen because: 1) in an intervention trying to integrate mental health services into primary care for post-conflict populations it is unethical to use a parallel design which prevents equal distribution of knowledge and skills; 2) the stepped wedge cluster design is more logistically feasible in the post-conflict setting of Northern Sri Lanka [15].

f. The authors stated: "Facilities are randomized using standardized randomization techniques to allocate facilities to the training sequence within the stepped wedge design. Patients and community representatives were not randomized. " Such statement needs more explanation and clarification

This has been clarified as follows in the manuscript (pages 8-9)

Facilities will be randomized using standardized randomization techniques to allocate facilities to the training sequence within the stepped wedge design. Patients and community representatives will not be randomized. The study team cannot be blinded to the allocation of facilities. However, the study psychiatrist will be masked to the randomization status of facilities from which patients originate. Patients will also be blinded to facility randomization status. [1]

g. Details are required for the training material that was used for training the primary care practitioners and public health professionals??? I suppose a detail is required for interventions among study groups!!

Additional detail outlining the training materials and intervention among study groups has been provided

Pages 11-12
Primary care practitioners will be trained using the mhGAP manual to recognize common mental health disorders using screening procedures, delivery brief psychosocial interventions (psychoeducation, Motivational Interviewing), and referral procedures for complex cases.
Page 13
Public health professionals will be trained using adapted written materials from the mhGAP training to recognize signs of common mental health issues, provide basic psychosocial care (education) and refer onwards if needed.
Page 14

Community representatives will be trained using adapted materials from the mhGAP training material to recognize signs of common mental illness, delivery basic psychosocial education and learn referral pathways and where to seek providers.

h. How the authors chose the patients to participate in the study ?? how they obtained the consent from patients to participate in the study ?? more clarification might be needed for such important issue!

Primary care practitioners selected patients using provided criteria, this has been clarified in the manuscript (page 12)

Further, before training, primary care practitioners will be asked to recruit up to eight patients using provided inclusion and exclusion criteria to have anxiety and/or depression diagnoses verified by a psychiatrist.

Consent procedures have been clarified (page 12)

Members of the research team will provide patients with information sheets and consent forms available in all 3 official languages (English, Tamil, Sinhala) and obtain informed consent.

i. On which basis the researchers chose three days of training, is it sufficient time for training such important issue??!

The 3-day training period was decided from a previous, similar study run in Sri Lanka, along with consultation from the WHO mhGAP training officer (Siriwardhana et al., 2016)

It would have been preferable to train for longer, however even training doctors over 3 days was a burden for them as we had to work with the MoH to find cover doctors which was difficult (there is only one doctor per clinic and we couldn't close the clinic)

j. Research team : who are the research team, what are the scientific background and experiences of such team ??

The research team consisted of senior academics from the UK, consultant psychiatrist and statistician from the Centers for Disease Control, consultant psychiatrist and community physician from Sri Lanka. All research assistants who collected the data were Sri Lankan and had experience in the research methods we used. All CVs for all research members were provided to the CDC in the initial funding application and approved.

k. Details are required for statistical methods including the variables and the program that would be used for statistics analysis, why they chose generalized linear regression to conduct such analysis?

Additional details on statistical methods have been included. A generalized linear regression was named as a possible statistical method by our statistician. The text has been amended as follows (pages 16-17)

The primary unit of analysis (i.e., where the intervention is carried out) is the healthcare facility, and data collected from the individual participants will include: sociodemographic information, conflict and displacement experience, mhGAP knowledge test scores, AMIQ stigma scores, and pre- and post-monitoring questionnaires. Descriptive statistics will be used to summarize observations using means, standard deviations, and proportions where appropriate.

An intention-to-treat analysis will be used to analyze the effect of the intervention. In the stepped wedge design, similar information will be collected repeatedly on individuals within facilities; we will therefore utilize statistical models that can adjust for repeat measures and clustering of participants within facilities (in the case of PHPs and CRs). In particular, for the analysis of mhGAP knowledge test scores and AMIQ stigma scores as outcomes of interest among participants, we plan to use generalized linear mixed models that include random intercepts for individuals (to account for repeat measurements), random intercepts for facilities (to account for the correlation among individuals within the same facility), and adjustments for covariates. Statistical analysis will be carefully conducted as repeated time measures over a 7-month time period will mean missing data points and shifted timelines due to external circumstances. Subgroup analysis will be conducted to understand differences between clusters and well as within clusters.

For qualitative data, interviews with primary care practitioners will continue until saturation has been met and thematic analysis will be used to analyse results.

Fidelity checks will be included in the study through the utilization of feedback forms after the end of each training session to investigate if any major concerns have arisen that could be addressed before subsequent sessions. Further, a process evaluation will be conducted to understand if the study adhered to protocol, and how the intervention was received by participants.

4. Missing information regarding the date of study conduction.

Additional information has been added to the introduction (page 5)

The second phase of the project was conducted between 2016-2021.

j. What are the limitations for such project ?

Limitations have been added to the beginning of the manuscript as per the journal editor's instructions (page 3)

Strengths and limitations of this study

First study to culturally adapt all World Health Organisation mhGAP written and video training materials
Seeks to demonstrate feasibility of using World Health Organisation mhGAP materials to train non-specialists
Multiple refresher sessions should increase knowledge retention among participants
Representative sample means findings could be generalized to other areas of Northern Sri Lanka
Limited to publicly available healthcare facilities as private facilities were excluded

5. References
Some references need to rewrite according to the style of the journal.

Thank you, references have been checked to ensure they are in the correct format

Response to Reviewer 2

Thank you very much for your comments and suggestions. I apologize if it wasn't clear, but this study was completed as of 30 April 2021 (it was submitted to the journal while the study was being conducted and was completed before reviewer comments were received).

The protocol has gone through a technical review from the funder (Centers for Disease Control and Prevention), and ethics reviews in the UK and Sri Lanka. I have addressed all comments below as best as I could, but am unable to revise some aspects of the protocol as the study has completed. Again, thank you for your careful review and thoughtful comments, it's greatly appreciated.

Aims
1.Was there a rationale for the percentage targets identified in the aims? For example, why was 50% chosen as a target for stigma reduction?

The sample size was calculated using a prevalence of depression or anxiety, 50% was chosen based on doubling previous prevalence rates in studies conducted in Mannar, Kilinochchi and Jaffna districts in Northern Province (Hussain et al, 2011; Siriwardhana et al, 2015). This information has been added to page 6

At least 50% reduction of mental health stigma measured using Attitudes to Mental Illness Questionnaire (AMIQ) in primary care practitioner, public health professional, and community representative participant groups at training, and 3- and 6-month follow-up. Prevalence rates were based on previous mental health studies conducted within the country [1,5].

I think stigma reduction is an interesting outcome and an important one to consider.

Thank you

2. The first aim appears to target increasing referrals to "specialist care for mental disorders." My understanding is that the purpose of this intervention is to provide mental health care in primary care settings, not refer to outside providers. Please provide clarity.

This aim was created to see if referrals for complex cases increased after primary care practitioners were trained to recognize mental health issues. They were trained only to treat

low level mental health issues and were asked to refer out cases that were more complex (for example psychosis). This aim was created by the funder and the original Principal Investigator (who passed away in April 2016). We were not able to completely meet this aim as we could only rely on feedback provided by the doctors pre- and post-training. We did supplement our findings by interviewing doctors after training to ask them if they felt more comfortable recognizing complex cases and referring them on. The aim has been amended to note "complex" cases to be clear (page 6)

30% increase in patients identified and treated for common mental health disorders, and referrals to specialist care for complex mental disorders by primary care practitioners. Measured at baseline using the pre/post-monitoring questionnaires on current practice with mental health patients and compared to post-training practice.

3. The second aim states that "qualitative interviews" will be used to determine concordance between PHC clinicians and psychiatrists. Please provide details regarding what this means. Is this a structured clinical interview of some sort?

Pre-training, doctors were asked to identify patients they thought had depression or anxiety. We then referred those patients to a local psychiatrist who verified if the diagnosis was correct or not (the psychiatrist reported back to the study team using a form). After training, we then interviewed the doctors using an open-ended topic guide. We asked them questions such as: "looking back on the patients you referred prior to training, would you do the same thing now?" "Do you feel more confident in identifying patients with mental health conditions now"?"

The second aim has now been amended to be clearer (page 6)

40% minimum concurrence between diagnoses of patients identified with mental disorders of interest by trained primary care practitioners and psychiatrists as compared to diagnosis of patients pre-training (baseline) measured using quantitative reporting forms from the consultant psychiatrist and in-depth, individual interviews with trained primary care practitioners to understand any changes in practice.

Study Design

1. May also note that using a cluster design can help to minimize contamination at sites

Thank you, we have added the following text to clarify both the limitations and strengths of using this design (page 9)

There is a theoretical possibility of contamination due to close professional or personal networks between participant groups and geographical location of facilities. However, this is not envisaged to exert significant influence on expected outcomes due to specifics of the extensive training, requirement to carry out intensive tasks, and low risk of contamination through informal discussions [1]. Potential contamination will be observed at all stages and taken into account at the end of the study. Further, an important strength of cluster sampling is that it helps mitigate the risk of treatment contamination between intervention and control groups.

2. How were the clinics randomized? Were they matched on any sort of criteria (e.g., size, location)?

A list of clinics was given to the study team by the Northern Province Ministry of Health. We used a random number generator to select the initial facilities. We only included public facilities: primary medical care units and divisional hospitals. We did not match on any criteria, but we did sample facilities proportionally to internal displacement (some districts were smaller but had larger internal displacement during the war, others were larger but had much less internal displacement). As internal displacement was a criterion of primary interest,

we sampled proportionally. Finally, we chose facilities district by district to minimize travel. This has been clarified as follows in the "Setting" section (page 9)

A list of public primary care facilities has been provided to the study team by the Northern Ministry of Health. A random number generator will be used to select 23 facilities across the 5 districts. Distribution of clusters will be allocated proportionally to the total numbers of internally displaced people in each district (districts with larger numbers were assigned more clusters). This will ensure adequate representation of conflict displacement and severity.

3. The details about timing is well-articulated and clear.

Thank you

Intervention:
1. Please provide brief details on the content of the training that the various participants received. I am unclear if they were trained on med prescribing, some sort of psychosocial intervention, etc. How was the training delivered – didactics, role plays, etc.?

Training consisted of delivery of written material, role plays, and videos. The emphasis was on psychosocial interventions as medications are not evenly available across districts. The "Structure and Delivery" section has been revised as follows to include this information (page 11)

Training will consist of delivery of written materials, role plays, and training videos. All training material was developed in collaboration with Sri Lankan researchers, local psychiatrists and community physicians. All material was translated, back translated, and piloted for cultural appropriateness prior to administration in the full project.

2. I appreciate the refresher courses; I think this is important to assist with retaining knowledge.

Thank you

3. Please provide brief details on the adaptation process of mhGAP. Was it simply translation? Were there any cultural/clinical/structural adaptations needed?

The mhGAP training materials were translated, back translated and then piloted for feasibility and cultural appropriateness. This was done in collaboration with Sri Lankan research, psychiatrists and community physicians. This has been added to the manuscript (page 11)

Almost all the instruments used in the study had been previously applied in epidemiological studies within Sri Lanka and are available in Tamil, Sinhala, and English languages. New instruments were translated, back translated and adapted for cultural appropriateness in collaboration with the Sri Lankan research team. Further, the entire set of measures was field tested in a pilot study prior to the full study commencing. This pilot study examined if measures were understood by participants, culturally acceptable, and to ensure translations were correct.

4. Were the measures adapted or validated in any way?

Almost all measures had been previously used in epidemiological studies in Sri Lanka and were available in Tamil, Sinhala and English. New instruments used were adapted for use in collaboration with the Sri Lankan research team and were field tested in a pilot study completed before the full study began. Please see above text found on page 11.

5. Please provide additional clarity about the roles of the various professionals engaged in the study. At this time, I am unclear about their functions:

Examples of each participant group has been added as follows

Primary care practitioners (page 12)

Primary care practitioners (general physicians practicing in selected facilities) will undergo one month of monitoring prior to training to establish a baseline.

Public health professionals (page 13)

Public health professionals (nurse attendants, midwives) in each selected facility will undergo one month of monitoring prior to training to establish a baseline.

Community representatives (page 14)

Community representatives (teachers, social workers) will undergo one month of monitoring prior to training to establish a baseline.

6. What is the role of public health professional in this study? Are they providing some sort of mental health intervention? What is the training that they receive?

Public health professionals (i.e., nurses, midwives) were trained to recognize symptoms of common mental health disorders (they are often the first set of people to speak to patients), deliver brief psychosocial interventions such as education, and trained to refer on if needed. This explanation has been added to the manuscript (page 13)

Public health professionals will be trained using adapted written materials from the mhGAP training to recognize signs of common mental health issues, provide basic psychosocial care (education) and refer onwards if needed.

7. What is the role of the community representative? Are they trying to increase referrals? Is the training mainly to reduce provider stigma? With whom do they engage – patients, providers?

Similar to public health professionals, community representatives (teachers, social workers) have lots of contact with the community so they were trained to recognize common symptoms of mental health disorders, deliver basic education and refer onwards if needed. They engage with both patients and providers. This information has been added to the manuscript (page 14)

Community representatives will be trained using adapted materials from the mhGAP training material to recognize signs of common mental illness, delivery basic psychosocial education and learn referral pathways and where to seek providers.

8. Please provide some details about how any safety issues (e.g., suicidality) will be managed.

A local psychiatrist was involved in the project from development to delivery. This psychiatrist was on call for any participants who required mental health help. This has been added to the "structure and delivery" portion of the manuscript (page 11)

To ensure safety of all participants, a local psychiatrist will be involved throughout the study to provide free mental health care if required.

9. What is "supervision" in the context of this intervention? Who provides supervision?

Supervision was provided by the research assistants who provided the training sessions to each group of participants. Supervision consisted of phone calls and in-person visits to see if and how participants were using training. This has been added to the manuscript (page 13)

Supervision at all time points will be provided by research assistants who previously delivered training sessions. Supervision will consist of phone calls or visits to the clinics to understand if and how training is being utilised by participants.

Analysis

1. If possible, including any information about measure validation would be very beneficial. Were there any fidelity checks/measures that can be included in the analyses?

Almost all measures had previously been validated in conflict-affected populations within Sri Lanka. Those that had not were field tested in a pilot as a validation measure. This has been added to the manuscript (page 11).

Almost all the instruments used in the study had been previously applied in epidemiological studies within Sri Lanka and are available in Tamil, Sinhala, and English languages. New instruments were translated, back translated and adapted for cultural appropriateness in collaboration with the Sri Lankan research team. Further, the entire set of measures was field tested in a pilot study prior to the full study commencing. This pilot study examined if measures were understood by participants, culturally acceptable, and to ensure translations were correct.

There were some fidelity checks included in the study – for example, feedback forms were given out at the end of every training to see if any major issues came up that could be revised for the next training. Further a process evaluation was run to check if the study adhered to protocol and how the intervention was received by participants. This has been added to the manuscript (page 17)

Fidelity checks will be included in the study through the utilization of feedback forms after the end of each training session to investigate if any major concerns have arisen that could be addressed before subsequent sessions. Further, a process evaluation will be conducted to understand if the study adhered to protocol, and how the intervention was received by participants.

2. Additional details about planned analyses are needed (e.g., reaching saturation, statistical approach, etc.).

Thank you, further information has been added to the manuscript after consultation with our statistician (pages 16-17).

The primary unit of analysis (i.e., where the intervention is carried out) is the healthcare facility, and data collected from the individual participants will include: sociodemographic information, conflict and displacement experience, mhGAP knowledge test scores, AMIQ stigma scores, and pre- and post-monitoring questionnaires. Descriptive statistics will be used to summarize observations using means, standard deviations, and proportions where appropriate.
An intention-to-treat analysis will be used to analyze the effect of the intervention. In the stepped wedge design, similar information will be collected repeatedly on individuals within facilities; we will therefore utilize statistical models that can adjust for repeat measures and clustering of participants within facilities (in the case of PHPs and CRs). In particular, for the analysis of mhGAP knowledge test scores and AMIQ stigma scores as outcomes of interest among participants, we plan to use generalized linear mixed models that include random intercepts for individuals (to account for repeat measurements), random intercepts for facilities (to account for the correlation among individuals within the same facility), and adjustments for covariates. Statistical analysis will be carefully conducted as repeated time measures over a 7-month time period will mean missing data points and shifted timelines due to external circumstances. Subgroup analysis will be conducted to understand differences between clusters and well as within clusters.
For qualitative data, interviews with primary care practitioners will continue until saturation has been met and thematic analysis will be used to analyse results.

Minor Notes
1. There are some minor typos/grammatical errors that may be improved with a close review.

Thank you, a close review has been completed to correct minor typos/grammatical errors throughout the manuscript.

2. Please provide the definition for the AMIQ the first time it is used in the text.

This has been changed in the aims portion of the manuscript (page 6).

At least 50% reduction of mental health stigma measured using Attitudes to Mental Illness Questionnaire (AMIQ) in primary care practitioner, public health professional, and community representative participant groups at training, and 3- and 6-month follow-up. Prevalence rates were based on previous mental health studies conducted within the country [1,5].

Response to Reviewer 3

Thank you very much for your comments and suggestions. I apologize if it wasn't clear, but this study was completed as of 30 April 2021 (it was submitted to the journal while the study was being conducted and was completed before reviewer comments were received).

The protocol has gone through a technical review from the funder (Centers for Disease Control and Prevention), and ethics reviews in the UK and Sri Lanka. I have addressed all comments below as best as I could, but am unable to revise some aspects of the protocol as the study has completed. Again, thank you for your careful review and thoughtful comments, it's greatly appreciated.

Overall (general) comments
1. The statistical analysis section needs more detail and would benefit from input by a suitably qualified medical statistician.

Thank you, additional detail has been provided about the statistical analysis from our statistician at the Centers for Disease Control and Prevention (pages 16-17).

The primary unit of analysis (i.e., where the intervention is carried out) is the healthcare facility, and data collected from the individual participants will include: sociodemographic information, conflict and displacement experience, mhGAP knowledge test scores, AMIQ stigma scores, and pre- and post-monitoring questionnaires. Descriptive statistics will be used to summarize observations using means, standard deviations, and proportions where appropriate.

An intention-to-treat analysis will be used to analyze the effect of the intervention. In the stepped wedge design, similar information will be collected repeatedly on individuals within facilities; we will therefore utilize statistical models that can adjust for repeat measures and clustering of participants within facilities (in the case of PHPs and CRs). In particular, for the analysis of mhGAP knowledge test scores and AMIQ stigma scores as outcomes of interest among participants, we plan to use generalized linear mixed models that include random intercepts for individuals (to account for repeat measurements), random intercepts for facilities (to account for the correlation among individuals within the same facility), and adjustments for covariates. Statistical analysis will be carefully conducted as repeated time measures over a 7-month time period will mean missing data points and shifted timelines due to external circumstances. Subgroup analysis will be conducted to understand differences between clusters and well as within clusters.

For qualitative data, interviews with primary care practitioners will continue until saturation has been met and thematic analysis will be used to analyse results.

2. From reading the full manuscript I can understand what your intended
overall aims are. However, a methods paper like this should have much more specific detail regarding the planned analysis – particularly in a clinical trial setting where pre-specification is so important. For example there is no explicit mention in the analysis section as to what variables are included in the analysis? Will inference be drawn from an adjusted or unadjusted model? How will you account for clustering (correlation) of individual patients within facilities?

Additional information has been added to clarify the statistical method as noted above.

3. The design of the study is a cluster randomised experiment and as such the primary unit of analysis is the facility. I think the authors should be more careful to highlight this, and be wary of individual level analysis as you may make inappropriate inferences from patient level analysis if the clustering is unaccounted for appropriately.

Thank you, additional information has been added to address this in the manuscript (page 8) Our project statistician is employed by the Centers for Disease Control and Prevention and has extensive experience with this kind of analysis.

This is a randomized clinical trial using a stepped wedge cluster design [14]. This design comprises a preliminary period where no clusters (healthcare facilities) are exposed to the intervention, then at regular intervals, one cluster is randomized to cross from control to intervention. This continues until all clusters have been exposed [14]. The stepped wedge design is increasingly utilized as a way to evaluate interventions that involve service delivery and involves sequential and random crossover of clusters (facilities) from control (delivery of standard care) to intervention (mhGAP training) until all clusters are exposed [14]. The stepped wedge design takes into account the logistical constraints of needing to sequentially roll out the intervention by training one facility at a time enabling us to understand how this intervention can be implemented in the future on a larger scale. This design was chosen because: 1) in an intervention trying to integrate mental health services into primary care for post-conflict populations it is unethical to use a parallel design which prevents equal distribution of knowledge and skills; 2) the stepped wedge cluster design is more logistically feasible in the post-conflict setting of Northern Sri Lanka [15].

4. The authors should specify a single primary outcome" rather than listing more than one. I understand the 3 choices reflects the different groups of participants that you hope the intervention impacts upon. But there should be a single primary outcome measure linked to the study design and sample size and remaining should be moved to secondary outcomes.

Thank you for this suggestion. While we appreciate the advice, these primary outcomes were used to guide the study (which was completed 30 April 2021) and we don't feel comfortable changing them now that the study has been completed.

5. The diagrams provided are helpful. They could be improved by being combined in to a single, simpler, schematic so a reader can see the overall picture.

Thank you, we have tried to combine all three diagrams into one, but it becomes too difficult to read so we have chosen to keep them separate.

6. There is no mention of key statistical challenges that this trial, like any other trial of a complex intervention, may encounter. These includes section on, but not limited to; missing data, multiple testing, adherence to intervention, planned sensitivity and/or subgroup analyses. Any further revisions of the paper should address these sections. Blinding (masking) is not discussed in the main body of the manuscript. Where blinding is possible it should be detailed. Where it is not possible then a discussion of the impact of this should be presented.

Thank you, additional information has been added (text is included under reviewer's previous point Overall Comments. (pages 16-17)

Blinding wasn't possible and this has been clarified in the manuscript (pages 8-9)

The study team cannot be blinded to the allocation of facilities. However, the study psychiatrist will be masked to the randomization status of facilities from which patients originate. Patients will also be blinded to facility randomization status. [1]

7. Contamination is important to consider in stepped wedge designs – why is there no discussions of potential contamination in the main manuscript?

Thank you, this information has been added (page 9).

There is a theoretical possibility of contamination due to close professional or personal networks between participant groups and geographical location of facilities. However, this is not envisaged to exert significant influence on expected outcomes due to specifics of the extensive training,

requirement to carry out intensive tasks, and low risk of contamination through informal discussions [1]. Potential contamination will be observed at all stages and taken into account at the end of the study. Further, an important strength of cluster sampling is that it helps mitigate the risk of treatment contamination between intervention and control groups.

8. Consider the CONSORT guidelines and relevant extension for stepped wedge designs (https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-018-3116-3). I would strongly recommend the authors consider this paper as it includes the checklist which the trial will have to report to, and as such the present manuscript should reflect on the areas of the checklist which are not suitably addressed.
Thank you, we have now included a CONSORT checklist

VERSION 2 – REVIEW

REVIEWER	Ameel Al Shawi University of Fallujah
REVIEW RETURNED	20-Sep-2021

GENERAL COMMENTS	Dear Sir / Madam I agree to accept the revised manuscript for publication

REVIEWER	Kimberly Hook Boston Medical Center, Psychiatry
REVIEW RETURNED	15-Sep-2021

GENERAL COMMENTS	Thank you for your revision and attention to the comments. After revision, I find that the manuscript more clearly details its aims and approach. A few minor points: •"Trained members of the research team will deliver training at each selected facility to minimize disruption of work at facilities." What training do these research team members have (masters, bachelor, doctoral level, mental health background, etc.)? •"Supervision at all time points will be provided by research assistants who previously delivered training sessions. Supervision will consist of phone calls or visits to the clinics to understand if and how training is being utilised by participants." Supervision does not seem to be clinical supervision, correct? Particularly for the physicians, it does not see that they are going to be evaluated on their accurate delivery of motivational interviewing or diagnostic skills? I would clarify for this point. At this time, all of my other comments have been addressed.

REVIEWER	James Griffin University of Warwick, Warwick Medical School, Clinical Trials Unit
REVIEW RETURNED	09-Sep-2021

GENERAL COMMENTS	Thank you for completing the recommended changes from the editor and reviewers. It has improved the manuscript a great deal. I think there is still scope for further specific details on the statistical analysis: -Which statistical program will be used? - Which subgroups will be investigated in the planned subgroup analysis?

	A pre-approved statistical analysis plan would also help support the eventual results publication and I urge the authors to complete a SAP with the assistance of their statistician, Thank you again for this important study protocol, and good luck with conducting the remainder of the study.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer 3

Thank you very much for your comments and suggestions. Please see our responses and revisions below.

1. Which statistical program will be used?
The following text has been included in the manuscript on page 14

Data analysis will be conducted using SAS version 9.4M7 [17].

2. Which subgroups will be investigated in the planned subgroup analysis?

The following text has been included in the manuscript on page 15

Planned subgroup analysis will investigate possible differences between clusters (for example, differences between PHPs across facilities).

3. A pre-approved statistical analysis plan would also help support the eventual results publication and I urge the authors to complete a SAP with the assistance of their statistician,

As part of our grant agreement, all statistical analysis will be handled by the statisticians at the Centers for Disease Control and Prevention, USA. They have created the analysis plan as presented in the manuscript based on their expertise and experience.

Response to Reviewer 2

Thank you very much for your comments and suggestions. Please see our responses and revisions below.

1. “Trained members of the research team will deliver training at each selected facility to minimize disruption of work at facilities.” What training do these research team members have (masters, bachelor, doctoral level, mental health background, etc.)?

The following text has been included in the manuscript on page 11

PCP trainers have Masters degrees in psychology and global mental health, and have successfully completed the WHO mhGAP train-the-trainers program.

The following text has been included in the manuscript on page 12

PHP trainers hold Bachelors’ degrees in psychology and have undergone 7 days training with senior members of the research team.

The following text has been included in the manuscript on page 13

CR trainers hold Bachelors’ degrees in psychology and have undergone 7 days training with senior members of the research team.

2. “Supervision at all time points will be provided by research assistants who previously delivered training sessions. Supervision will consist of phone calls or visits to the clinics to understand if and how training is being utilised by participants.” Supervision does not seem to be clinical supervision, correct? Particularly for the physicians, it does not see that they are going to be evaluated on their accurate delivery of motivational interviewing or diagnostic skills? I would clarify for this point.

This is correct, supervision was to understand if and how training was utilised by participants, this was not clinical supervision.

The following text has been included in the manuscript on page 11-12

Non-clinical supervision at all time points will be provided by research assistants who previously delivered training sessions. Supervision will consist of phone calls or visits to the clinics to understand if and how training is being utilised by participants. Supervision is meant to understand if and how trained participants utilise mhGAP, accurate delivery of interviewing or diagnostic skills were not evaluated.

The following text has been included in the manuscript on page 13

Non-clinical supervision at all time points will be provided by research assistants who delivered previous training sessions.

The following text has been included in the manuscript on page 14

Non-clinical supervision at all time points will be provided by research assistants who delivered previous training sessions.