Author’s response to reviews

Title: A study protocol for testing the feasibility of a randomized stepped wedge cluster design to investigate a Community Health Intervention through Musical Engagement (CHIME) for perinatal mental health in The Gambia

Authors:

Katie Rose Sanfilippo (ksanf001@gold.ac.uk)

Bonnie McConnell (Bree.McConnell@anu.edu.au)

Victoria Cornelius (v.cornelius@imperial.ac.uk)

Buba Darboe (darboejula@hotmail.com)

Hajara Huma (hajarahuma24@gmail.com)

Malick Gaye (malickgaye60@yahoo.com)

Paul Ramchandani (pr441@cam.ac.uk)

Hassoum Ceesay (hceesay@gmail.com)

Vivette Glover (v.glover@imperial.ac.uk)

Ian Cross (ic108@cam.ac.uk)

Lauren Stewart (l.stewart@gold.ac.uk)

Version: 1 Date: 16 Aug 2019

Author’s response to reviews:

Below I have provided the same text which is included in the review response document attached.

We would like to thank both reviewers for their thoughtful and helpful comments. Below we detail our responses to each comment (in italics) as well as highlighting any substantial changes to the manuscript, both by including new text in this document (with page numbers) and highlighting its position in our revised document. We have also changed the order of the aims (keeping the wording the same) and, in line with this, the order of the analysis plan. We have also taken the opportunity to make some small re-wording edits to improve flow and readability (all highlighted in green).

Reviewer 1:
As point 1-3 are related we respond to all three below:

1) The study is described as a feasibility study for a stepped wedge trial. However, I wasn’t clear about the process. It reads as if there is going to be an intervention and control group at the same facility, but I presume that is not going to be the case because of the risk of contamination. I expected that the randomisation would be at the clinic level and would be in terms of which clinics were randomised to get the intervention in step 1 vs step 2. If this is the case, please can you make that clearer? If this is not the case, then I am not sure it is a stepped wedge trial.

2) Related to my first query, assuming it is a stepped wedge design, I understand that all pregnant women will be recruited at the baseline of the study and then will have repeated measures of their mental health, thus allowing the clinics to act as their own controls. If so, how will you manage the fact that the women will be further along in the pregnancy gestation by the time they get the intervention? (if they are in the clinics randomised to the 2nd step).

3) Please can the investigators justify why they selected a stepped wedge design rather than a cluster RCT?

We have now amended the document in the Methods and Trial Schedule sections to be clearer about the trial design and our rationale for choosing this design over a cluster design. We have also clarified how contamination will be avoided and include an updated figure (in place of Table 1 and 2) to better illustrate the trial design.

We will be testing the feasibility of a stepped wedge cluster design, which differs from a parallel arm cluster design in that all clinics involved in the study receive the intervention (Hemming et al., 2018). Advantages over a parallel arm cluster trial include the requirement of a smaller sample size due to the availability of a within group comparison and prevention of potential disappointment for health clinics who are not randomized into the intervention.

(Methods/Design section, p 6, lines 101-105)

This trial design involves a sequential crossover of clusters whereby each cluster (antenatal clinic) receives the control condition followed by the intervention condition. The four chosen clinics will be randomized to two sequences of a 12-week phase. A separate cohort of participants will be recruited to the control group and the intervention (singing) group so each cohort will be recruited around 4-6 months into their pregnancy. The 12 week phase for both the control and intervention cohort will include data collection at week 1 (‘baseline’) and week 7 (‘post’) after either group-singing (intervention) or standard care (control) as well as at week 11 (‘follow up’), four weeks after the intervention finished. Contamination will be
avoided because all data for the control group will be collected before the intervention groups start at each clinic. See Figure 1 for a schematic for the study.

(Trial Schedule section, p 13, lines 252-263)

Figure 1. Schematic for the stepped wedge cluster design

(inserted in Trial Schedule section, p 13, line 263)

4) The investigators propose to include all pregnant women, but some will not have any depressive symptoms and so it will not be possible to see any improvement. Is the intervention being conceptualised as preventing poorer mental health as well as alleviating current symptoms? If so this should be made explicit and this should be reflected in relevant design aspects. If the intention is to have a universal intervention that does not require screening of depressive symptoms (which could have benefits in terms of personnel training + acceptability and avoiding stigma) then this rationale should be mentioned.

Thank you for this helpful comment. The goal is to deliver the intervention on the community level including all women rather than just those with a higher level of symptoms. We therefore explain our rationale for doing so more explicitly within the Intervention section of the document as below:

The intervention will be delivered on the community level, meaning that it will include women with a range of anxiety and depression symptoms. Our primary aim is to reduce symptoms in those experiencing them (whether these are at a high level or a medium or low level). We anticipate this may also help to reduce their symptoms into the postnatal period. By including those with low and high levels of symptoms, rather than screening and including only those with high levels of symptoms we will aim to avoid stigma and increase acceptability.

(Intervention section, p 7, lines 124 – 130)

5) Relatedly, what about women who have severe depressive symptoms? Shouldn't there be an element of stepped care whereby more severely unwell women are identified at the beginning and fast-tracked for psychiatric nurse review (to consider risk management, potential prescription of antidepressants, potential admission, need for more intense follow-up and monitoring)? Is it ethical to keep severely depressed women in the groups without additional care when the evidence base indicates that this type of intervention is most likely to benefit women with mild-moderate depressive symptoms, but not severe?

Thank you for raising this very important consideration. Within our protocol we have tried to make clear where women will be referred on if they have been identified as experiencing
suicidal thoughts or behaviours and/or domestic violence. This can be seen as points 2 and 3 under the Assessment and Management of Risks section. We follow the same procedure of referrals for women who have been identified as experiencing severe symptoms. However, we do realise this was not made clear in our manuscript. As our RAs are both trained psychiatric nurses, both are well-equipped to identify those women and are embedded in the mental health care system in The Gambia. While the resources available to women are extremely limited, as there are only two trained psychiatrists in the entire country, we have tried our best to make sure women who have been identified as needing more support are referred on to the best care available. Also, as the access to mental health care is so limited in the Gambia, we would suggest that including women with high levels of symptoms in the study can be justified, since there is no evidence to suggest that this type of intervention could worsen symptoms, and may, on the contrary provide some level of support. The only women with severe symptoms who were excluded were those presenting with symptoms of psychosis. We have now made this clear by adding another point (4) under the Assessment and Management of Risks section.

4. Women throughout the study will be monitored by the RAs, both trained psychiatric nurses. If the RAs feel that at any point a woman’s score indicates a high level of symptoms and/or the women reveal that they are particularly struggling, the RAs will refer the woman on to the Community Mental Health Team (CMHT) for further management. If the CMTH deems it appropriate they may then refer her on to the psychiatric team.

(Assessment and Management of Risk section, p 14-15, lines 297 – 302)

6) The investigators face a difficult challenge because of the inclusion of women speaking two different languages. Without a validation study, it will not be possible to combine scores from these different groups, given that cut-offs tend to vary when translated into different languages/adapted across cultural groups. Even with a validation study it is going to be difficult. Can this feasibility study be used to investigate and inform how the investigators might need to handle this issue in the main trial?

This is an extremely helpful comment and we thank you for your input. We agree that collapsing across the different languages could be potentially problematic. Therefore, we will examine the distributions of scores in the different language groups to see to what extent item scores and overall distributions are similar across the two versions. This will also inform the planning and design of a future definitive trial. Our plan to understand this potential difference is now explained in the Analysis section of the manuscript. We have also taken out the section of the analysis plan that discusses comparing different cut offs (“We will also calculate the proportion of women who are above threshold for potential depression and anxiety using published validated cut-offs across a range of African countries for comparison.”) as we felt this type of analysis was not applicable to this feasibility trail as there are no validated cut offs for these measurements tools for use in The Gambia. We have
added a sentence to state how we plan to investigate differences between the two language groups in the Analysis section.

We will also examine the distributions of scores in the different language groups to see to what extent item scores and overall distributions differ or are similar.

(Analysis section, p 15-16, lines 320 – 321)

7) To what extent is there a need for an intervention for distress that may well be transient? Is it worth trying to also measure level of disability which would help to make the case for the health system relevance of the intervention?

Symptoms of high anxiety and depression are persistent in a high proportion of women, and there is a large body of evidence showing strong associations between high anxiety and/or depression (measured at one or more time points) and a higher risk of adverse outcomes for mothers and their infants in the postnatal period. We have now been clearer about our rationale for the intervention in the Background section of the manuscript.

Stress, anxiety and depression in pregnancy affect not only the mother but can also have long-term adverse effects on her child via biological mechanisms in utero (Glover, O’Donnell, O’Connor, & Fisher, 2018). Along with the impact on the mother and her developing infant, antenatal depression and anxiety are the most common predictors of postnatal depression (Milgrom et al., 2008; Robertson, Grace, Wallington, & Stewart, 2004). Postnatal depression can reduce her ability to provide sensitive and responsive caregiving, and this can also impair child development (Prenoveau et al., 2017).

(Background section, p 3 , lines 35 - 41)

We agree that a measure of disability, quality of life and the related cost effectiveness analyses could be a valuable addition and will consider this in a future full-scale trial.

8) Can the investigators include an evaluation of the reliability of the application of the fidelity measures?

During the workshops we have set out clear goals and aims for the intervention sessions including the structure and the content of the songs. From this we have created a checklist to allow us to ascertain fidelity of the intervention sessions. We now make this clear with a small addition in the description of the intervention in the Intervention section.

Over the course of the six-week intervention period a research assistant will observe and video and audio record each of the four settings twice (the first and fourth week) to ascertain, using a checklist, the extent to which the sessions conform to our articulated goals.
We have also explained in more detail how we will test for reliability in the Analysis section. Since both RAs will separately watch the videos and complete the checklist, we will measure and report inter-rater consistency.

To determine if the intervention is deliverable we will record the number of sessions that the Kanyeleng groups delivered, the duration of each session and will perform a qualitative evaluation, using the video and audio recordings, to determine intervention fidelity at the four sites. Both RAs will watch the video and audio recordings of the 1st and 4th group-singing sessions at each clinic and complete a check list to determine if the necessary elements - as outlined in the training workshops - were included in the intervention. Reliability of the fidelity measure will be ascertained by measuring inter-rater consistency.

9) If women are not literate, would there need to be a witness to confirm that the information was explained as per the information sheet?

To ensure that all participants received the relevant information, the RAs read out the information sheet verbatim. It was not possible to have a witness for each consent procedure, but that the RA’s were trained, monitored and were provided with peer support and supervision throughout. A sentence has been added to clarify this in the Recruitment section of the manuscript.

The information sheet will be read out in their native language verbatim to ensure participants’ ability to give informed consent.

10) In terms of the concern that women may link adverse events to participation in the intervention, perhaps it would be wise to provide clear information about the nature of the intervention at baseline to try to avoid any misunderstandings. Involving MoH to convince women after the event may back-fire (and the specifics of any SAE/AE needs to be considered in order to decide relatedness or not, so it is not possible to have a global statement on this).

Thank you for your advice regarding this point. The protocol that we developed in the case of SAE or AE was formed on the advice from a panel of local practitioners in the area. Within The Gambian context, this type of response is usually what is recommended. However, we also understand the importance of clearly explaining the nature of the intervention and any
possible risks at baseline. We have made this clearer in the protocol in the Assessment and Management of Risks section.

We will mitigate this association being made, firstly, by clearly explaining the nature of the intervention and any possible risks to the women when they are recruited into the study.

(Assessment and Management of risks section, p 13-14, lines 274-276)

11) Is there any religious connotation to singing in this setting and will this mean that any women are systematically excluded (if they do not endorse that religious faith)? Could this be monitored in the feasibility trial?

The Gambia is predominately Muslim with a significant Christian minority. The song repertoire being used in the study is not religious and people of all faiths take part. In community settings, it is common for Muslims and Christians to attend each other's celebrations and sing and dance together. That said, some conservative Muslims choose not participate in this kind of singing. This remains a relatively small minority in the Gambia, but it is possible that some women may decline to participate in the study for religious reasons. To help understand this, and potentially other factors that might lead to certain groups being systematically excluded, we will make sure to collect the reasons why some women might choose not to participate. This had been made clear in the Recruitment section.

If participants do not choose to take part we will record their reason, if it is given, to help understand why women might not want or be able to participate.

(Recruitment Section, p 12, lines 229-230)

Reviewer 2:

1) The abstract describes the topic of interest and its relevance and methodology well. The outcome measures are written in abbreviated form in the abstract and I'd suggest to write them in full and use abbreviated name versions later.

Thank you. This has now been amended in the abstract.

We will assess symptoms of anxiety and depression using the Edinburgh Postnatal Depression Scale (EPDS) and the Self Reporting Questionnaire (SRQ-20).

(Abstract, p 2, lines 18-20)

2) In the background, it may be useful to add a few references specific to perinatal music therapy and perinatal music-based interventions as these are increasingly used in the perinatal period.
We have now added a line and some references to point to work specifically using music interventions during the perinatal period.

Recent studies have found that music and its use specifically during the perinatal period can decrease women’s stress levels, depressive symptoms and increase women’s attachment to their infant (11-14).

(Background, p 4, lines 59-61)

3) In addition to using EPDS and SRQ-20, it would be helpful to consider adding a brief quality of life measure to outcome measures.

We agree that an additional quality of life measurement would be helpful. At this stage of the trial, we are unable to add any more measures but this feedback will be extremely helpful when we decide on what measures to include when planning our for the next stage.

4) Under methodology, it may be beneficial to add a sentence explaining stepped wedge design so that the readers can better understand the recruitment and intervention allocation.

Thank you for your helpful feedback. This is similar to feedback given by reviewer one and a significant change has been made to make our design clearer. Below we have copied our response to reviewer 1’s comments.

We have now amended the document in the Methods and Trial schedule sections to be clearer about the trial design and our rationale for choosing this design over a cluster design. We have also clarified how contamination will be avoided and include an updated figure (in place of Table 1 and 2) to better illustrate the trial design.

We will be testing the feasibility of a stepped wedge cluster design, which differs from a parallel arm cluster design in that all clinics involved in the study receive the intervention (Hemming et al., 2018). Advantages over a parallel arm cluster trial include the requirement of a smaller sample size due to the availability of a within group comparison and prevention of potential disappointment for health clinics who are not randomized into the intervention.

(Methods/Design section, p 6, lines 101-105)

This trial design involves a sequential crossover of clusters whereby each cluster (antenatal clinic) receives the control condition followed by the intervention condition. The four chosen clinics will be randomized to two sequences of a 12-week phase. A separate cohort of participants will be recruited to the control group and the intervention (singing) group so each cohort will be recruited around 4-6 months into their pregnancy. The 12 week phase for both the control and intervention cohort will include data collection at week 1 (‘baseline’) and week 7 (‘post’) after either group singing (intervention) or standard care (control) as well as a at week 11 (‘follow up’), four weeks after the intervention finished. Contamination will be...
avoided because all data for the control group will be collected before the intervention groups start at each clinic. See Figure 1 for a schematic for the study.

(Trial Schedule section, p 13, lines 252-263)

Figure 1. Schematic for the stepped wedge cluster design

(inserted in Trial Schedule section, p 13, line 263)