Author’s response to reviews

Title: A nurse-delivered mental health intervention for obstetric fistula patients in Tanzania: Results of a pilot randomized controlled trial

Authors:
Melissa Watt (melissa.watt@duke.edu)
Mary Mosha (maryanfort@yahoo.com)
Alyssa Platt (alyssa.platt@duke.edu)
Kathleen Sikkema (kathleen.sikkema@duke.edu)
Sarah Wilson (sarah.wilson@duke.edu)
Elizabeth Turner (elizabeth.turner@duke.edu)
Gileard Masenga (drgmasenga@yahoo.com)

Version: 1 Date: 27 Jun 2017

Author’s response to reviews:

May 15, 2017

Ref: PAFS-D-16-00112 “A nurse-delivered mental health intervention for obstetric fistula patients in Tanzania: Results of a pilot randomized controlled trial”

Dear Dr. Hughes,

My co-authors and I appreciate your review of our manuscript, and your invitation for us to resubmit this paper. Below you will find our responses to the reviewer’s comments, with page numbers referring to the revised manuscript. We have sincerely appreciated the thoughtfulness of the reviewers and your recognition of the importance of this topic and the suitability to your journal. The insights provided by the reviewers have helped us to think more deeply and broadly about this work and have greatly enhanced the quality and clarity of the manuscript.
We look forward to your review of this manuscript and to the possibility of having our work published in Pilot and Feasibility Studies.

Reviewer #1

METHODS

1. Page 4, line 22 - You say a sample size of 60 was chosen as adequate based on the reference from Browne et al. (1995). Did you perform calculations taking into account effect size to arrive at this number? If so, I think it would be best to report all these values, and if not, you should elaborate on the justification of 60 as an adequate sample size.

Our sample size was primarily driven by an adequate sample to examine the feasibility and acceptability of a pilot RCT with two conditions. We did, however, conduct a power calculation at the outset, and have now included those details. (Page 4, Lines 10-14)

"A pilot RCT was conducted between March 2014 and June 2016.[23, 24] A target sample size of 60 was chosen as adequate to examine the feasibility and acceptability of a pilot RCT with two conditions.[25] We estimated the power to detect a difference in 3-month primary outcomes for the sample size of 60 women (30 per group) would be 77%, based on ANCOVA and assuming within-person correlation between baseline and 3-month depression scores of 0.7 and a pooled standard deviation (SD) of 12. Such a scenario would correspond to an effect size of 0.5."

2. Page 5, line 26 - While you cannot blind participants or those delivering the intervention, there are other study staff which could be blinded to the assigned conditions, such as personnel doing the allocation, data analyses, etc. Additionally, if women in both conditions completed identical surveys, those delivering them could also be blinded so as not to introduce any subtle biases when asking the questions. I'm wondering if you considered this and/or did this. If not, I think it should be a point brought up as a potential limitation.
The statistician who conducted the data analysis was blinded to intervention condition. The intervention condition was only revealed at the conclusion of analysis.

We had considered blinding the study staff involved in data collection, but it was not feasible in the context of this study. Given resource limitations, we had personnel who were performing multiple roles (e.g., both tracking participants through the study, including their intervention exposure, and also conducting the survey interviews). We have now added a statement about this to the limitations. (Page 15, Lines 15-17)

The study staff who were involved in data collection were not blinded to intervention condition, due to a limited number of individuals playing multiple study roles. This may have impacted reporting of outcomes and differential follow-up rates, and should be reconsidered for future studies.

3. Page 6, line 6 - When describing the intervention, add in how long the sessions typically were, especially given you mentioned some participants were not satisfied with the length of the sessions in the results section.

We now clarify that the intervention sessions were approximately 60 minutes each.

4. Page 6, line 51 - Were the patients' obstetric history self-reported? And if so, were there answered verified at all using medical records?

Obstetric history was self reported. Given that this was typically patients’ first time at the study hospital, this information would not have been included in the medical record (except as self-reported patient history). We have now added clarification that the obstetric history was self-reported.
5. Page 8, line 56 - Did you have any feasibility criteria for recruitment rates?

We have now added the following about our feasibility of recruitment rates: “and aimed to recruit an average of 4 patients per month.” Our recruitment rate did not meet this criteria, which we now report in our results section: “Enrolment was slower than expected, at a rate of approximately 2.5 per month.”

6. Page 8, line 15 - You mentioned intervention feasibility is defined by the number of sessions attended. Was there a benchmark percentage of attendance levels that you chose to indicate the intervention being successful or not?

We now clarify that we used 80% as a benchmark.

7. Page 8, line 27 - Provide some more detail about the participant feedback. In particular, what questions were asked? If possible, consider providing a copy of the questionnaire outline as a supplementary file.

We have now provided details about the 15 questions that were asked to assess the acceptability of the intervention. (Page 8, Line 24, to Page 9, Line 7)

"Acceptability of the intervention was assessed from participants in the intervention condition with 15 questions at the post-treatment survey. We asked five questions about satisfaction with various domains, each with structured response options (“Would you take part in this study again?”, “How satisfied were you with the intervention?”, “How satisfied were you with the intervention?”, “What did you think about the number of sessions?”, “What did you think about the time spent at each session?”). We also asked four open-ended questions, covering details about what the participant liked, disliked, would what want to see changed, and any general feedback on the overall structure and content of the intervention. Finally, we asked a question about the perceived usefulness of each of the six sessions (responses ranging from not useful to very useful)."
RESULTS:

8. Page 10, line 60 - Expand on what exactly you defined as "appropriate" regarding the intervention for patients. What sorts of responses did the patients provide that helped you come to this conclusion?

We have removed the word “appropriate” and reordered that section of the text to be more clear. (Page 11, Lines 19-22)

"The clinical notes and debriefing meetings revealed that, overall, the patients connected well with the interventionist, and the intervention content allowed them to forthright in sharing their life experiences, emotions, and concerns."

DISCUSSION:

9. Did you adapt the study protocol at all based on the experience of conducting the pilot study or the feedback from patients? For example, it was mentioned that follow-up attrition rates varied based on things such as PTSD and surgery success. Given these results, you may consider proposing ways to increase follow-up completion rates for subgroups of patients who may be most at-risk for dropout.

At this point, we do not have funding to conduct a full RCT, although we hope this might be possible in the future. We have included a sentence in the discussion that this issue of follow-up completion rates should be considered in a future trial. (Page 16, Lines 8-10)

"Additional efforts should be made to ensure retention of women over time, particularly considering women with high levels of distress or whose surgeries were not successful and differential retention by intervention condition."
Reviewer #2

1. Page 3, line 11 - I recommend adjusting the language 'when prolonged obstructed labour is not relieved by cesarean delivery' because it does not adequately address the time component. Prolonged obstructed labor may be relieved by cesarean delivery but not before adequate tissue damage has been done to result in the development of a fistula.

Thank you for this clarification. We have now reworded: (Page 3, Line 4)

"Women develop obstetric fistula due to prolonged obstructed labour, or when a perforation is made during a caesarean section."

2. Page 3, line 40 - The language in this sentence may be too optimistic given the varying definitions of treatment success and levels of persistent incontinence following surgery.

We have now modified to temper the language and optimism of this sentence: (Page 3, Line 18)

"Surgical repair may heal or improve an obstetric fistula, and many countries such as Tanzania have established free fistula repair programs.[18]"

3. Page 4, line 35-40 - it may be helpful to describe in more detail the average number of fistula patients at any one time given the potential for cross-contamination of intervention effect due to what has been described in the randomization protocol. Furthermore, the literature on women and girls affected by fistula shows that patients often have not met others with the condition prior to accessing treatment at a hospital or camp and that the social support provided in such a context by others having undergone similar experiences is quite significant.
We have now provided this detail: “During the course of the study, there were anywhere from zero to five patients admitted for fistula repair surgery any one time.”

We agree that cross-contamination is very likely in our study design, and we have noted this in the discussion, both in the interpretation of the findings (“women were recruited and individually randomized from a single fistula ward, and contamination across conditions was likely common.”) and also in our recommendation for future studies (“Given the potential for contamination in an individually-randomized trial, a cluster RCT design should be considered.”)

4. Page 5, line 38 - could you contextualize the post data collection point in the treatment process for fistula, i.e., did this occur after catheter removal and woman's experience of surgical outcome?

Yes, we consistently conducted the post data collection after the catheter removal and approximately 24 hours prior to discharge, so that woman’s experience of surgical outcome was clear.

5. Page 6, line 16 - to be completely clear, the community health nurse was not hospital-based staff that was conducting any other clinical care with patients aside from the intervention over the course of the study?

That is correct. She was not providing clinical care and only interacted with the patients in the context of this study.
6. Page 6, lines 18-27 - this section would be improved by inclusion of greater detail on the intervention targets.

We have published another paper that describes the theoretical underpinnings of the intervention, which is referenced in this section. However, we agree that additional details about the intervention would improve the study. We have now added the following: (Page 6, lines 12-17)

"The manualized curriculum was developed through extensive qualitative work with a range of stakeholders, and based on theories of cognitive behavioral therapy[28] and coping models.[29] By helping patients to reframe their experience with a fistula and develop coping skills to deal with the physical and social impact of the fistula, the intervention aimed to reduce patient distress, improve self esteem and optimism, and therefore support successful community reintegration following repair."

7. Page 6, lines 26-29 - the paper would be improved by inclusion of the theory of change behind the intervention.

Please see above.

8. Page 6, line 33 - the surveys at time points baseline, post-surgery and 3-month follow-up were all identical?

Yes, the surveys were identical.

9. Page 6, line 40 - please insert who the survey was administered by. If administered by the community health nurse, this will need to be described as a potential limitation due to social desirability bias.
The surveys were delivered by a research staff member who was not involved in intervention delivery. As noted in response to reviewer 1 (comment #2), the staff member was not blinded to intervention, which we have now added as a study limitation.

10. Page 6-7, lines 56-40 - could you specify whether the alpha coefficients are representing populations similar to your patient population?

The alpha coefficients were calculated from the data collected in this study.

11. Page 8, line 55-56 - could you please provide rationale for adjusting by length of time with fistula and indicate whether this was modeled as a linear effect? I question whether the relationship of time with fistula and the outcomes is in fact linear in shape. I also wonder whether this value was mean centered for interpretability.

The main purpose of inclusion of this variable was as a control for confounding. Since we also expect this variable to be prognostic of the outcome, we hoped to improve the precision of our treatment effect estimates by including it. It was standardized (i.e. mean-centered and divided by the sample standard deviation) for interpretability. We have now added that detail in the analysis section. It was modeled as a linear effect.

12. Page 9, lines 4-7 - I may not be adequately knowledgeable to comment on this component, but question whether the mental health scales included in the study can purport to establish meeting diagnostic criteria versus simply measuring the varied symptomatology. Perhaps this could be established earlier at the introduction of the measures? Additionally, it would be useful to establish that this was done using the baseline survey data (vs. baseline clinician assessment?).

This is an excellent point, and you are correct that we shouldn’t imply that women have been diagnosed with a mental disorder given their scores on the mental health screeners. The cut-off
scores have been established in the broader literature via rigorous studies of scale validity. We have now adjusted some of the wording throughout, using language such as “symptomatology score indicative of a diagnosis”.

13. Page 9, lines 4-47 - Could you please comment on the adequacy of the sample sizes used to establish the functional versus the dysfunctional distributions?

Sample sizes for depression were N=65 (functional) and N=40 (dysfunctional)
Sample sizes for anxiety were N=35 (functional) and N=25 (dysfunctional)
Sample sizes for PTSD were N=49 (functional) and N=44 (dysfunctional)

These sample sizes are relatively small but have the advantage of being sourced from context-appropriate populations rather than larger samples from less representative Western nations and populations. All but one of the above samples is greater than the N=30 suggested to ensure that the sample mean is drawn from a normal distribution according to the Central Limit Theorem, increasing the likelihood that the mean is close to the true central tendency of the population. Larger standard deviations drawn from smaller samples also help ensure that our estimates are conservative (i.e. biased towards the null).

We have now included the sample sizes for the functional distributions in the methods, where we describe the clinical significance analysis.

Ideally, we would have a large independent sample drawn from comparable population of women in Western Africa; however, since no such data are available, we chose to exchange precision for validity in this case.
14. Page 10, lines 13-15 - Differentiating follow-up by intervention status is of interest in this section.

We have now added this detail: “with slightly higher follow-up in the intervention condition (76.7% vs. 70%).”

15. Page 10, lines 25-39 - The implication of the differential follow-up on the study findings should be included in the discussion section.

Please see the response to Review #1, comment 9.

16. Page 10, line 42 - Again I wonder about the use of the term diagnostic criteria, and anticipate this comment will be resolved in response to my earlier question.

As noted to the comment above, I have now updated this language throughout.

17. Page 12, lines 33-36 - I find it is difficult to put only the percentages here given the small ns - could this be updated to the style used on page 11, line 27 instead?

18. Discussion general - Please address the 73% retention rate in this section.

This has now been done – please see above.
19. Discussion general - Please address the differing scales of the mental health measures in interpreting magnitude of intervention effect.

The following sentences have been added to the discussion section (page 15, lines 2-4):

“Since the mental health outcome measures used for regression analysis have different numerical scales, interpretation of the findings should consider that differences observed in estimated means will have different clinical meanings for each scale.”

20. Page 14, lines 56-60 - This sentence could be expanded to include development of the intervention and based additionally on formative research with key stakeholders, in addition to psychological theory.

We have now added that detail to this sentence.

21. Table 1 - Distance from home to KCMC is presented in varying levels of precision across intervention vs. control group.

We have now updated Table 1 to be consistent in our precision, with one decimal point.

22. Table 1 - it would be informative to include the definition of successful surgery in a note under the table.

We now include the following footnote:
3Surgical success was assessed by physician and defined as a closed fistula with no more than stress incontinence and no need for repeat surgery.

23. Table 3 - it would be informative to include the definitions of recovered and reliably improved/deteriorated in a note under the table.

Definitions are now included at the bottom of the table.

24. Table 3 - please add the revised overall ns for the analyses of the 3-month follow-up data on the same lines as they are in the first half of the table.

Sample sizes are included on the same lines as in the first half of the table and have also been added to footnotes.