PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Comparing Four Service Delivery Models for Adolescent Girls and Young Women through the "Girl Power" study: Protocol for a Multisite Quasi-experimental Cohort Study
AUTHORS	Rosenberg, Nora; Pettifor, Audrey; Myers, Laura; Phanga, Twambilile; Marcus, Rebecca; Bhushan, Nivedita; Madlingozi, Nomtha; Vansia, Dhrutika; Masters, Avril; Maseko, Bertha; Mtwisha, Lulu; Kachigamba, Annie; Tang, Jennifer; Hosseinipour, Mina C.; Bekker, Linda Gail

VERSION 1 – REVIEW

REVIEWER	Ghyslain Mombo-Ngoma Centre de Recherches Médicales de Lambaréné (CERMEL) Gabon
REVIEW RETURNED	05-Sep-2017

GENERAL COMMENTS	The study protocol is well written with the study question and objectives clearly expressed as well as the study outcomes of interest. The authors have described very well the limitations of the protocol. It is an ongoing study though closer to its end, it deserves to be published

REVIEWER	Matthew F Chersich With RHI, University of the Witwatersrand, South Africa
REVIEW RETURNED	06-Sep-2017

GENERAL COMMENTS	The protocol sums a very important study, by a seasoned group of investigators. The methods have several notable strengths. The intervention delivery will require a considerably amount of time and efforts of the group, who are congratulated for being willing to commit their energies to the study. The beauty of the study is, however, not captured well in the paper, which does not do the study justice, not reflect the level of thought that has clearly gone into the study. I understand that problems with the study design cannot be corrected now, and I am not requesting that occur. But I feel that if the investigators wish to publish the protocol, they must give a more critical appraisal of the methods used in the study. Acknowledge the weaknesses of the study, and how these have been mitigated, why options that would have incurred fewer limitations were not selected. In my view, a protocol paper should aim to give enough info that the study could be 'exactly' replicated by another study.

	There are major gaps in the info given in this paper, such that replication would be difficult. The rationale used for the decisions made in the design is only given in a few instances, but often not in enough detail. Major revisions 1. The authors give conflicting info on whether previous studies have assessed a similar intervention. P2, line 3 'have not been compared'. 'have not been rigorously evaluated' this implies it has been done and evaluated, but not in a rigorous manner. 'there are few examples' implies there are some, but no reference is given for this. 'never been tested'. 'is a major focus of donors' Regardless of what is true about this issue, the reader is told this several times more than is necessary. 2. The strengths and limitations section of the paper is very poor, and then the limitations are repeated later in the paper. Please restrict the strengths and limitations section to the study design and methods. Currently it includes aims and other things. 3. Present the outcome measures used in a figure, this forces the investigators to use very specific terms and note the exact primary outcome, source of data for this etc. There is a lot of vagueness about these measures, for example 'set of primary outcomes' and 'primary outcome of interest'. Are women using contraception at baseline included in the denominator for the endline contraception measure (or only those without the outcome of interest), and is it contraceptive use or unmet need etc. How are discrepancies resolved between reported outcomes and outcomes noted in clinic records etc. 4. The analysis methods lack detail. What assumptions will you check b4 using GEE? How has clustering been taken into account. And what level of clustering will you use and why: country, clinic, small group in Model 3, 4? 5. Sample size calculation does not provide enough detail to be replicable, what test was used in the comparison, clustering effect? Give the actual equation used Limitations 1. Generalisability concerns around including only Western Cape sites in SA, a province very distinct from the rest of the country. And what is the predominant race group enrolled, how does that influence generalisability? 2. Selection biases are possible, and may vary by group, a major limitation I think. The first 3 groups may enrol 'worried well', who differ systematically from other young women in the area. The fourth group may enrol young women who appreciate the opportunity for monthly income. Differential incentives to participate apply to different study groups. 3. Differential levels of loss to follow up across the groups is likely. The authors do not state how missing data will be treated, and whether they consider this missing at random etc.? 4. I am not convinced that contamination is not possible at all, perhaps note the distance between clinics to make this claim more plausible 5. Note if there is a possibility that having extra staff in the intervention facilities will in any possible way influence the quality of data collection on services provided (the outcome measure). Additional staff in the intervention clinics, whose work is being monitored over a year will have quite some incentive to collect data on all contraceptives in 'existing clinic records' dispensed, for example, while in the Model 1, such incentives do 6. Will pregnancy incidence be measured? I am surprised that the authors do not believe this outcome is possible.

	Please substantiate that claim with citing pregnancy incidence rates in this population. Data on this in the 2016 SA DHS would clearly suggest otherwise, in my view (about 15% of 15-19 years olds already pregnant once)? Malawi rates are even higher, amongst the highest in the world 7. Limitations of analysis: a. Given the step-wise escalation of intervention one would like to know in analysis whether there was indeed a stepwise simple dose-type effect, or a threshold effect, where a single or double intervention don't work, but only all 3. This may not be possible to assess given the design 8. The Figure 1 is not good. There are inconsistencies between the figure and the texts 'clinic navigation is indicated as happening in all Malawi clinics in figure, but not in text in model 1. Different terms are used in the Malawi side of the figure to the SA side, for what seems to be the same thing: 'discuss SRH needs in SA', but 'determine SRH in Malawi'. If the interventions do differ then use different terms, if it is a matter of consistency of language errors, then correct pls. 'Young' is sometimes defined, but othertimes not, and defined as <29 even. 9. Ethical considerations a. I wonder about naming the clinics, if one lands up being very poorly functioning in the study, you then have no way of maintaining confidentiality about the clinic, for example. Similar principles apply to naming of clinics as to naming of individuals in a study? b. The concerns about enrolling adolescents is not discussed in detail, in fact the authors do not note if assent is taken from girls. How were the rules in SA about minister approval circumvented? 10. LARCs appear to be treated the same as other contraceptives, despite their very different effectiveness, especially in this age group. I would consider this outcome measure (contraception use) to be a weak indicator of the distal thing it measures (risk for pregnancy), given that it only partially reflects the risk, conflating very different things (LARC and OC, for example). Note the strengths of design, e.g. 1. Multi-site major strength, raises generalisability in particular. Minor revisions 1. Is it a patient held card 2. No as-treated analysis? 3. The measure of outcomes (service use) is not stated clearly. 4. State if condoms free in Malawi (p6 last 3rd of page) 5. Give names of tools used to measures alcohol and depression etc. 6. 'non-integrated separate location with adults' is not easy to understand, rephrase pls. 7. State if the protocol has been registered 8. State if pregnancy incidence, STI incidence is measured 9. Why are 19 year olds placed in the adolescent group, reference this classification 10. When existing evidence is presented, pls give study design so reader can see strength of other relevant studies (Malawi study on top of page 6) 11. Page 6 'a cash transfer program has never been operationalised'?? I doubt that is true, never say never, and never with a behavioural intervention? 12. Someone would argue that strictly speaking you are doing a cluster randomised trial, explain which features of a CRT are present in your design and which not, and thus why you say it is not a CRT 13. You name the clinics in Figure 1, but so no need to mention them on page 7.

	Optional changes What alternative designs were possible? Consider stating the amount of funding required for such a study, reporting the study protocol is a chance to provide useful info for the reader, info that a reader wishes to know. Am interested to know the rationale for not providing this?

REVIEWER	Matthew F Chersich With RHI, University of the Witwatersrand, South Africa
REVIEW RETURNED	06-Sep-2017

GENERAL COMMENTS	The protocol sums a very important study, by a seasoned group of investigators. The methods have several notable strengths. The intervention delivery will require a considerably amount of time and efforts of the group, who are congratulated for being willing to commit their energies to the study. The beauty of the study is, however, not captured well in the paper, which does not do the study justice, not reflect the level of thought that has clearly gone into the study. I understand that problems with the study design cannot be corrected now, and I am not requesting that occur. But I feel that if the investigators wish to publish the protocol, they must give a more critical appraisal of the methods used in the study. Acknowledge the weaknesses of the study, and how these have been mitigated, why options that would have incurred fewer limitations were not selected. In my view, a protocol paper should aim to give enough info that the study could be 'exactly' replicated by another study. There are major gaps in the info given in this paper, such that replication would be difficult. The rationale used for the decisions made in the design is only given in a few instances, but often not in enough detail. Major revisions 1. The authors give conflicting info on whether previous studies have assessed a similar intervention. P2, line 3 'have not been compared'. 'have not been rigorously evaluated' this implies it has been done and evaluated, but not in a rigorous manner. 'there are few examples' implies there are some, but no reference is given for this. 'never been tested'. 'is a major focus of donors' Regardless of what is true about this issue, the reader is told this several times more than is necessary. 2. The strengths and limitations section of the paper is very poor, and then the limitations are repeated later in the paper. Please restrict the strengths and limitations section to the study design and methods. Currently it includes aims and other things. 3. Present the outcome measures used in a figure, this forces the investigators to use very specific terms and note the exact primary outcome, source of data for this etc. There is a lot of vagueness about these measures, for example 'set of primary outcomes' and 'primary outcome of interest'. Are women using contraception at baseline included in the denominator for the endline contraception measure (or only those without the outcome of interest), and is it contraceptive use or unmet need etc. How are discrepancies resolved between reported outcomes and outcomes noted in clinic records etc.

	4. The analysis methods lack detail. What assumptions will you check b4 using GEE? How has clustering been taken into account. And what level of clustering will you use and why: country, clinic, small group in Model 3, 4? 5. Sample size calculation does not provide enough detail to be replicable, what test was used in the comparison, clustering effect? Give the actual equation used Limitations 1. Generalisability concerns around including only Western Cape sites in SA, a province very distinct from the rest of the country. And what is the predominant race group enrolled, how does that influence generalisability? 2. Selection biases are possible, and may vary by group, a major limitation I think. The first 3 groups may enrol 'worried well', who differ systematically from other young women in the area. The fourth group may enrol young women who appreciate the opportunity for monthly income. Differential incentives to participate apply to different study groups. 3. Differential levels of loss to follow up across the groups is likely. The authors do not state how missing data will be treated, and whether they consider this missing at random etc.? 4. I am not convinced that contamination is not possible at all, perhaps note the distance between clinics to make this claim more plausible 5. Note if there is a possibility that having extra staff in the intervention facilities will in any possible way influence the quality of data collection on services provided (the outcome measure). Additional staff in the intervention clinics, whose work is being monitored over a year will have quite some incentive to collect data on all contraceptives in 'existing clinic records' dispensed, for example, while in the Model 1, such incentives do 6. Will pregnancy incidence be measured? I am surprised that the authors do not believe this outcome is possible. Please substantiate that claim with citing pregnancy incidence rates in this population. Data on this in the 2016 SA DHS would clearly suggest otherwise, in my view (about 15% of 15-19 years olds already pregnant once)? Malawi rates are even higher, amongst the highest in the world 7. Limitations of analysis: a. Given the step-wise escalation of intervention one would like to know in analysis whether there was indeed a stepwise simple dose-type effect, or a threshold effect, where a single or double intervention don't work, but only all 3. This may not be possible to assess given the design 8. The Figure 1 is not good. There are inconsistencies between the figure and the texts 'clinic navigation is indicated as happening in all Malawi clinics in figure, but not in text in model 1. Different terms are used in the Malawi side of the figure to the SA side, for what seems to be the same thing: 'discuss SRH needs in SA', but 'determine SRH in Malawi'. If the interventions do differ then use different terms, if it is a matter of consistency of language errors, then correct pls.

	Young' is sometimes defined, but othertimes not, and defined as <29 even. 9. Ethical considerations a. I wonder about naming the clinics, if one lands up being very poorly functioning in the study, you then have no way of maintaining confidentiality about the clinic, for example. Similar principles apply to naming of clinics as to naming of individuals in a study? b. The concerns about enrolling adolescents is not discussed in detail, in fact the authors do not note if assent is taken from girls. How were the rules in SA about minister approval circumvented? 10. LARCs appear to be treated the same as other contraceptives, despite their very different effectiveness, especially in this age group. I would consider this outcome measure (contraception use) to be a weak indicator of the distal thing it measures (risk for pregnancy), given that it only partially reflects the risk, conflating very different things (LARC and OC, for example). Note the strengths of design, e.g. 1. Multi-site major strength, raises generalisability in particular. Minor revisions 1. Is it a patient held card 2. No as-treated analysis? 3. The measure of outcomes (service use) is not stated clearly. 4. State if condoms free in Malawi (p6 last 3rd of page) 5. Give names of tools used to measures alcohol and depression etc. 6. 'non-integrated separate location with adults' is not easy to understand, rephrase pls. 7. State if the protocol has been registered 8. State if pregnancy incidence, STI incidence is measured 9. Why are 19 year olds placed in the adolescent group, reference this classification 10. When existing evidence is presented, pls give study design so reader can see strength of other relevant studies (Malawi study on top of page 6) 11. Page 6 'a cash transfer program has never been operationalised'?? I doubt that is true, never say never, and never with a behavioural intervention? 12. Someone would argue that strictly speaking you are doing a cluster randomised trial, explain which features of a CRT are present in your design and which not, and thus why you say it is not a CRT 13. You name the clinics in Figure 1, but so no need to mention them on page 7. Optional changes What alternative designs were possible? Consider stating the amount of funding required for such a study, reporting the study protocol is a chance to provide useful info for the reader, info that a reader wishes to know. Am interested to know the rationale for not providing this?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Ghyslain Mombo-Ngoma
Institution and Country: Centre de Recherches Médicales de Lambaréné (CERMEL), Gabon
Competing Interests: None declared

Comment: The study protocol is well written with the study question and objectives clearly expressed as well as the study outcomes of interest. The authors have described very well the limitations of the protocol. It is an ongoing study though closer to its end, it deserves to be published

Response: We thank the reviewer for taking the time to review our manuscript and provide these comments.

Reviewer: 2
Reviewer Name: Matthew F Chersich
Institution and Country: With RHI, University of the Witwatersrand, South Africa Competing Interests: None declared

Comment: The protocol sums a very important study, by a seasoned group of investigators. The methods have several notable strengths. The intervention delivery will require a considerably amount of time and efforts of the group, who are congratulated for being willing to commit their energies to the study.
The beauty of the study is, however, not captured well in the paper, which does not do the study justice, not reflect the level of thought that has clearly gone into the study. I understand that problems with the study design cannot be corrected now, and I am not requesting that occur. But I feel that if the investigators wish to publish the protocol, they must give a more critical appraisal of the methods used in the study. Acknowledge the weaknesses of the study, and how these have been mitigated, why options that would have incurred fewer limitations were not selected. In my view, a protocol paper should aim to give enough info that the study could be 'exactly' replicated by another study. There are major gaps in the info given in this paper, such that replication would be difficult. The rationale used for the decisions made in the design is only given in a few instances, but often not in enough detail.

Response: We thank the reviewer for this thoughtful and critical reading of our manuscript. We have addressed his concerns in point-by-point responses below.

Major revisions
1. The authors give conflicting info on whether previous studies have assessed a similar intervention. P2, line 3 'have not been compared'. 'have not been rigorously evaluated' this implies it has been done and evaluated, but not in a rigorous manner. 'there are few examples' implies there are some, but no reference is given for this. 'is a major focus of donors' Regardless of what is true about this issue, the reader is told this several times more than is necessary.

Response: We have revised the language in the introduction for clarity. We have deleted one of the two instances where donor attention was mentioned.

2. The strengths and limitations section of the paper is very poor, and then the limitations are repeated later in the paper. Please restrict the strengths and limitations section to the study design and methods. Currently it includes aims and other things.

Response: Thank you for this feedback. This section has been completely rewritten to focus on design and methodologic strengths and weaknesses. We deleted the original limitations section.

3. Present the outcome measures used in a figure, this forces the investigators to use very specific terms and note the exact primary outcome, source of data for this etc. There is a lot of vagueness about these measures, for example 'set of primary outcomes' and 'primary outcome of interest'. Are women using contraception at baseline included in the denominator for the endline contraception measure (or only those without the outcome of interest), and is it contraceptive use or unmet need etc. How are discrepancies resolved between reported outcomes and outcomes noted in clinic records etc.

Response: We have removed the vague terminology and provided more detail on our outcome measures. As suggested we have created a table of outcome measures that indicate the data source, numerator, and denominator.

4. The analysis methods lack detail. What assumptions will you check b4 using GEE? How has clustering been taken into account. And what level of clustering will you use and why: country, clinic, small group in Model 3, 4?

Response: Analyses will be run separately in each country. The clinic is the primary exposure of interest and therefore cannot be a clustering variable. Clustering will account for multiple observations within each individual in each quarter.

5. Sample size calculation does not provide enough detail to be replicable, what test was used in the comparison, clustering effect?

Response: We have provided more detail on how our sample size was calculated. Our power calculations address any two-way comparison between arms at any time point.

Limitations
1. Generalisability concerns around including only Western Cape sites in SA, a province very distinct from the rest of the country. And what is the predominant race group enrolled, how does that influence generalisability?

Response: We only had funding to implement the study in one set of sites in South Africa. We have described the racial composition of our clinics and discussed the generalizability concerns. We clarify: "In South Africa, the four sites are in the Klipfontein and Mitchell's Plain areas in the Western Cape, which serve primarily black and coloured populations." We have also discussed this in the strengths and limitations section.

2. Selection biases are possible, and may vary by group, a major limitation I think. The first 3 groups may enrol 'worried well', who differ systematically from other young women in the area. The fourth group may enrol young women who appreciate the opportunity for monthly income. Differential incentives to participate apply to different study groups.

3. Differential levels of loss to follow-up across the groups is likely. The authors do not state how missing data will be treated, and whether they consider this missing at random etc.?

Response: We have discussed differential recruitment and differential loss-to-follow-up as potential limitations, and noted that even though our four models are different our recruitment and retention procedures were the same.

4. I am not convinced that contamination is not possible at all, perhaps note the distance between clinics to make this claim more plausible

Response: We have added the following: “In South Africa, biometric identification is being used to ensure the same people do not enroll in more than one site. In Malawi, all sites are at least 7km apart.” We have also mentioned contamination as a potential limitation.

5. Note if there is a possibility that having extra staff in the intervention facilities will in any possible way influence the quality of data collection on services provided (the outcome measure). Additional staff in the intervention clinics, whose work is being monitored over a year will have quite some incentive to collect data on all contraceptives in ‘existing clinic records’ dispensed, for example, while in the Model 1, such incentives do [not exist].

Response: We thank the reviewer for raising this critical issue. We have added the following: “In all clinics, study staff will examine clinical records to obtain missed records and ensure consistent ascertainment.” We have added the concern of differential data ascertainment to the limitations.

6. Will pregnancy incidence be measured? I am surprised that the authors do not believe this outcome is possible. Please substantiate that claim with citing pregnancy incidence rates in this population. Data on this in the 2016 SA DHS would clearly suggest otherwise, in my view (about 15% of 15-19 years olds already pregnant once)? Malawi rates are even higher, amongst the highest in the world 7.

Response: We are measuring pregnancy incidence in an exploratory fashion in Malawi only. Even with very high incidence rates of 15 per 100 person years, we would need nearly perfect retention and large differences between sites to have sufficient power to observe statistically significant differences in pregnancy.

a. Limitations of analysis: Given the step-wise escalation of intervention one would like to know in analysis whether there was indeed a stepwise simple dose-type effect, or a threshold effect, where a single or double intervention don’t work, but only all 3. This may not be possible to assess given the design

Response: Ideally, we would have had a fifth clinic in each country with YFHS + cash transfer (but no empowerment sessions). This would allow us to observe whether the cash + clinic together had a synergistic effect greater than either of these two elements alone.

8. The Figure 1 is not good. There are inconsistencies between the figure and the texts ‘clinic navigation is indicated as happening in all Malawi clinics in figure, but not in text in model 1.
I believe the reviewer is referring to Table 1, not Figure 1. Thank you for noticing this contradiction.

Response: We have modified the table accordingly.

Different terms are used in the Malawi side of the figure to the SA side, for what seems to be the same thing: ‘discuss SRH needs in SA’, but ‘determine SRH in Malawi’. If the interventions do differ then use different terms, if it is a matter of consistency of language errors, then correct pls. ‘Young’ is sometimes defined, but other times not, and defined as <29 even.
We have edited the figure for consistency of language. The remaining differences in the Malawi and South African sides of the table are intentional. We have defined young one time as <29 years. Most of the Malawi staff were in their 20s—they were young enough to relate participants and mature enough to maintain confidentiality.

9. Ethical considerations
a. I wonder about naming the clinics, if one lands up being very poorly functioning in the study, you then have no way of maintaining confidentiality about the clinic, for example. Similar principles apply to naming of clinics as to naming of individuals in a study?

Response: Thank you for catching this oversight! We had intended to delete these. We will write to the editor regarding deletion from the original submission to preserve anonymity and deleted from the resubmission.

b. The concerns about enrolling adolescents is not discussed in detail, in fact the authors do not note if assent is taken from girls. How were the rules in SA about minister approval circumvented?

Response: We have developed this section of our manuscript. We now state “We requested that minors 15-17 years be able to consent for themselves because they are able to receive all of these services without parental consent. In a study designed to reduce barriers to care-seeking, obtaining parental consent could pose an undue barrier.”

10. LARCs appear to be treated the same as other contraceptives, despite their very different effectiveness, especially in this age group. I would consider this outcome measure (contraception use) to be a weak indicator of the distal thing it measures (risk for pregnancy), given that it only partially reflects the risk, conflating very different things (LARC and OC, for example).

Response: Thanks for this important observation. It is true that different contraceptive methods have very different contraceptive efficacy and effectiveness. In addition to observing whether or not contraception is being used, we will report on the method mix.

Note the strengths of design, e.g. 1.Multi-site major strength, raises generalisability in particular.
We have added this design strength.

Minor revisions
1. Is it a patient held card

Response: It is not a patient-held card. This has been added. It is housed at the clinic. This clarification has been added to the data collection and management tab.

2. No as-treated analysis?

Response: As-treated analyses re-analyze data based on actual exposure, rather than intended exposure. In our analysis, the primary exposure is the clinic one is assigned to. We do not have a way of determining if participants in one clinic received services in another.

3. The measure of outcomes (service use) is not stated clearly.

Response: These have been clarified in our outcomes table.

4. State if condoms free in Malawi (p6 last 3rd of page)

Response: Condoms are free in Malawi. This has been added.

5. Give names of tools used to measures alcohol and depression etc.

Response: We used the 10-item Center for Epidemiologic Studies Depression scale, modified conflict tactic scale to assess intimate partner violence, and the National Institute on Alcohol Abuse and Alcoholism's alcohol screening assessment.

6.'non-integrated separate location with adults' is not easy to understand, rephrase pls.

Response: We have rephrased this clause in the abstract. "AGYW can receive family planning, HIV testing and counseling (HTC), and STI syndromic management in 3 separate locations with 3 separate queues with the general population. No youth-friendly spaces, clinic modifications or trainings are offered."

7.State if the protocol has been registered

Response: The protocol has not been registered.

8.State if pregnancy incidence, STI incidence is measured

Response: We have added that pregnancy incidence is being measured in Malawi.

9. Why are 19 year olds placed in the adolescent group, reference this classification

Response: The World Health Organization considers persons 10-19 years to be adolescents. This reference has been added. http://www.who.int/topics/adolescent_health/en/

10.When existing evidence is presented, pls give study design so reader can see strength of other relevant studies (Malawi study on top of page 6)

Response: We have indicated the requested information in this instance, as well as several others.

11.Page 6 'a cash transfer program has never been operationalised'?? I doubt that is true, never say never, and never with a behavioural intervention?

Response: We have rephrased and no longer say "never."

12.Someone would argue that strictly speaking you are doing a cluster randomised trial, explain which features of a CRT are present in your design and which not, and thus why you say it is not a CRT

Response: A cluster RCT would have multiple clinics in each arm in each country. Our study only has one clinic in each country in each arm. We have mentioned this in the strengths and limitations section.

13. You name the clinics in Figure 1, but so no need to mention them on page 7.

Response: We have removed all indications of clinic names. As indicated, this was an oversight. We have also written to the editor to discuss removal of these clinic names from the original submission.

Optional changes
What alternative designs were possible?

Response: There are multiple alternative designs and we wish to remain focused on the one we selected.

Consider stating the amount of funding required for such a study, reporting the study protocol is a chance to provide useful info for the reader, info that a reader wishes to know. Am interested to know the rationale for not providing this?

Response: We have complied with the journal's funding statement which does not suggest inclusion of the study budget.

VERSION 2 – REVIEW

REVIEWER	Matthew F Chersich WRHI, SA
REVIEW RETURNED	13-Oct-2017

GENERAL COMMENTS	Many important changes have been made, but overall I feel the responses we more superficial than detailed, and the authors did not fully engage with some of the comments. This pertains especially to comments 2, 3, 4, 5 and 6a (the response does not give any indication of how any stepwise effects might be measured). The limitations section remains weak in my view. The statement that 'efforts are being made to ensure comparable… ' does not tell the reader about the likely direction and size of these biases, nor that the investigator have considered the specific biases that apply to this study, and how to minimise them specifically. The sentence on bias in this study (bullet 2) would equally apply to a survey of 10 postmen. The authors state they have: 'added concern of differential data ascertainment to the limitations section'. Again, 'we will ensure comparable data ascertainment across sites' does not explore this threat in sufficient detail. The use of several 'primary outcomes' is interesting, as opposed to a single outcome, please comment on that decision. For pregnancy incidence, imagine a simple measure of saying 15% of adolescents have had a pregnancy at 12 months in Group 1 (the estimate given by the authors), and only 6% in group 4, you would require 203 evaluable adolescents in each arm to detect a difference: stata command is sampsi 0.15 0.06, power(0.8). The actual power calculation for the study assumes a 0% loss to follow up, yet the authors say this must be taken into account in a pregnancy sample size calculation. Query 9b was also only responded to in part: how were the rules about ministerial approval for adolescent assent waived?

VERSION 2 – AUTHOR RESPONSE

Reviewer's Comments to Author:
Many important changes have been made, but overall I feel the responses we more superficial than detailed, and the authors did not fully engage with some of the comments. This pertains especially to comments 2, 3, 4, 5 and 6a (the response does not give any indication of how any stepwise effects might be measured).

>We thank the reviewer for his thoughtful consideration of our study's limitations. We ask the reviewer to note that a detailed discussion of the limitations is not part of journal requirements. Nonetheless, we have added a discussion section in order to address the remaining concerns. The original comments (2, 3, 4, 5, and 6a, as well as 6 and 9) are included below for ease of reference. We have interspersed the original and new comments and our original and new responses. We trust that this more detailed discussion of our limitations is satisfactory.

2. Original comment: Selection biases are possible, and may vary by group, a major limitation I think. The first 3 groups may enrol 'worried well', who differ systematically from other young women in the area. The fourth group may enrol young women who appreciate the opportunity for monthly income. Differential incentives to participate apply to different study groups.

>New response: We agree that there are possibilities for selection bias in our study, although do not necessarily agree that models 2 and 3 would recruit the worried well. We have added the following sentence to the discussion section: "each model may preferentially recruit persons who are interested in the services at that clinic. For example, the absence of services in model 1 may lead to recruitment of persons who do not need any services and the presence of a cash transfer in model 4 may lead to recruitment of persons in need of cash. We will explore whether baseline behavioral and socio-economic characteristics differ by model, and if so conduct adjusted analyses as needed."

3. Original comment: Differential levels of loss to follow-up across the groups is likely. The authors do not state how missing data will be treated, and whether they consider this missing at random etc.?

Original response (Note it responded to 2 and 3): We have discussed differential recruitment and differential loss-to-follow-up as potential limitations, and noted that even though our four models are different our recruitment and retention procedures were the same.

>New response: We have added the following sentences to the discussion: "Second, it is possible that retention will be differential across arms. Different retention is precisely what we are trying to measure in clinical outcomes, but is problematic with respect to behavioral survey outcomes. To mitigate this risk, identical research incentives will be offered in all models at the three behavioral survey visits. We will explore the magnitude and nature of loss by clinic and use multiple imputation techniques and sensitivity analyses to address the loss that does occur."

4. Original comment I am not convinced that contamination is not possible at all, perhaps note the distance between clinics to make this claim more plausible

Original response: We have added the following: "In South Africa, biometric identification is being used to ensure the same people do not enroll in more than one site. In Malawi, all sites are at least 7km apart." We have also mentioned contamination as a potential limitation.

>New response: Enrollment in more than one site is impossible in South Africa due to the biometric identification system. It is possible in the Malawi site and we now describe this as a limitation: “Finally, enrollment in more than one model is possible in the Malawi sites, which do not have biometric identification. If such contamination were to occur at a large scale, behavioral survey results would likely be biased towards the null, as response options from the same participant at multiple clinics would be similar.”

5. Original comment: Note if there is a possibility that having extra staff in the intervention facilities will in any possible way influence the quality of data collection on services provided (the outcome measure). Additional staff in the intervention clinics, whose work is being monitored over a year will have quite some incentive to collect data on all contraceptives in ‘existing clinic records’ dispensed, for example, while in the Model 1, such incentives do [not exist].

Original response: We thank the reviewer for raising this critical issue. We have added the following: “In all clinics, study staff will examine clinical records to obtain missed records and ensure consistent ascertainment.” We have added the concern of differential data ascertainment to the limitations.

New comment: The authors state they have: ‘added concern of differential data ascertainment to the limitations section’. Again, ‘we will ensure comparable data ascertainment across sites’ does not explore this threat in sufficient detail.

>New response: We have added the following to the limitations section. “Differential ascertainment of clinical outcomes is a third potential source of bias: clinical staff trained in models 2, 3, and 4 may capture services more consistently than staff in clinic 1: apparent differences in service uptake could in fact be differences in data capture. Observing whether clinical records are consistent with self-report will allow us to explore this potential bias.”

6. Original comment: Will pregnancy incidence be measured? I am surprised that the authors do not believe this outcome is possible. Please substantiate that claim with citing pregnancy incidence rates in this population. Data on this in the 2016 SA DHS would clearly suggest otherwise, in my view (about 15% of 15-19 years olds already pregnant once)? Malawi rates are even higher, amongst the highest in the world 7.

Original response: We are measuring pregnancy incidence in an exploratory fashion in Malawi only. Even with very high incidence rates of 15 per 100 person years, we would need nearly perfect retention and large differences between sites to have sufficient power to observe statistically significant differences in pregnancy.

New comment: For pregnancy incidence, imagine a simple measure of saying 15% of adolescents have had a pregnancy at 12 months in Group 1 (the estimate given by the authors), and only 6% in group 4, you would require 203 evaluable adolescents in each arm to detect a difference: stata command is sampsi 0.15 0.06, power(0.8). The actual power calculation for the study assumes a 0% loss to follow up, yet the authors say this must be taken into account in a pregnancy sample size calculation.

>New response: We appreciate the thought experiment and believe the author’s sample size calculation actually reinforces our position. A reduction of pregnancy incidence from 15% to 6% would represent a very large impact that we are not confident that we can achieve. Additionally, we would only want to only focus on those AGYW who did not want to get pregnant, further limiting our sample size and making the primary outcome not generalizable to the full study population. We are indeed planning a paper on the impact of the interventions on pregnancy incidence in the Malawi cohort, but are not planning to change our primary outcomes at this late stage.

Should we see evidence of favorable trends, we may pursue a larger study that includes pregnancy incidence as a biomarker outcome.

a. Limitations of analysis: Given the step-wise escalation of intervention one would like to know in analysis whether there was indeed a stepwise simple dose-type effect, or a threshold effect, where a single or double intervention don't work, but only all 3. This may not be possible to assess given the design.

Original response: Ideally, we would have had a fifth clinic in each country with YFHS + cash transfer (but no empowerment sessions). This would allow us to observe whether the cash + clinic together had a synergistic effect greater than either of these two elements alone.

>New response: I am not sure if I understand the question. In primary analyses, we will compare participants in models 2, 3, and 4 to those in model 1. However, we also plan to rotate the reference groups and compare models 2 and 3 to one another and models 3 and 4 to one another. These comparisons will help us understand if the addition of each component had an incremental impact. If there are more sophisticated techniques that the reviewer wishes to describe, we are open to considering them!

The limitations section remains weak in my view. The statement that 'efforts are being made to ensure comparable… ' does not tell the reader about the likely direction and size of these biases, nor that the investigator have considered the specific biases that apply to this study, and how to minimise them specifically. The sentence on bias in this study (bullet 2) would equally apply to a survey of 10 postmen.

>New response: The editor reminded us that the bullets are supposed to be one only sentence long, rather than a detailed discussion of limitations. We hope the new limitations section described above addresses the reviewer's concerns.

The use of several 'primary outcomes' is interesting, as opposed to a single outcome, please comment on that decision.

>New response: The core purpose of our study is to assess the influence of integrated interventions on a range of outcomes. Although we realize this is not traditional, we believe it is more meaningful than a single-outcome study.

Original comment: 9b. The concerns about enrolling adolescents is not discussed in detail, in fact the authors do not note if assent is taken from girls. How were the rules in SA about minister approval circumvented?
Original response: We have developed this section of our manuscript. We now state "We requested that minors 15-17 years be able to consent for themselves because they are able to receive all of these services without parental consent. In a study designed to reduce barriers to care-seeking, obtaining parental consent could pose an undue barrier."

New comment: Query 9b was also only responded to in part: how were the rules about ministerial approval for adolescent assent waived?

>New response: The justification described above was elaborated on and presented to the UCT Committee on Research Ethics and they approved it. We now state this in the manuscript. We believe this decision did not place 15-17 year old young women at any increased risk and that it resulted in a more representative sample, as we did not restrict ourselves to those who could obtain parental permission.

VERSION 3 – REVIEW

REVIEWER	Matthew F Chersich WRHI
REVIEW RETURNED	08-Nov-2017

GENERAL COMMENTS	Paper much improved, well done.