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)	Measuring the prevalence and impact of poor menstrual hygiene management: a quantitative survey of schoolgirls in rural Uganda.
AUTHORS	Hennegan, Julie; Dolan, Catherine; Wu, Maryalice; Scott, Linda; Montgomery, Paul

VERSION 1 - REVIEW

REVIEWER	Penelope A Phillips-Howard Liverpool School of Tropical Medicine, UK
REVIEW RETURNED	01-Jun-2016

GENERAL COMMENTS	The study aims to quantify the main components of menstrual hygiene management (MHM), aggregate respondent's answers to measure if 'adequate' MHM influences some key outcome indicators. Authors note the small sample size, self-reported information, interviewer translation of questions while conducting the face-to-face interviews (with reverse translation for data entry per question), and the cross-sectional nature of the study, all of which affect interpretation of findings. Of note, the survey interviewers could have encountered a desirability effect as they were the same staff that had provided the intervention of AFRIpads in the former trial (it is unclear if they are still distributing, or these are very old pads). Nevertheless, despite these limitations, the approach toward quantifying information is an important contribution toward this hitherto neglected topic. The results are concerning, particularly the implication that provision of reusable pads did not significantly contribute towards adequate MHM. The implication that AFRIpads do not improve MHM, and that girls who reported adequate MHM did not appear to have significantly improved outcome measures is difficult to explain, and may in part reflect the small sample size, and inability to control for confounding. If the provision of AFRIpads truly did not improve MHM, this has implications for programmes wishing to scale-up provision. The authors suggest disposable pads may be more valuable, as they do not depend on washing and drying, however disposal and cost of such items needs to also be considered in the implications. We note no mention is made of menstrual cups, although they require minimum WASH facilities, are reusable, and cost-effective. The authors can help readers better understand and interpret the findings, by strengthening parts of the manuscript, with attention to the following suggestions. Methods First line in methods is rather odd, it adds nothing as the specific details of a correct study should be in the methods. More detail is needed about the trial. It would be useful to know if girls provided with AFRIpads were truly randomised in the trial, it

	states the trial was quasi-randomised, are the girls in the control arm different from those not receiving AFRIpads? Was school the unit of randomisation? Did the authors perform sample size calculations for which the small sample was chosen, particularly to allow stratification by age and school? The study population is 205 girls, I can see no reason for mentioning the non-menstruating girls who were non-eligible. It was not clear to the reviewer if provision of AFRIpads was ongoing, and if the trial was completed. It is difficult to define if it is part of the trial, or is a stand-alone cross sectional survey, after closure. It is relevant to know how long girls were provided AFRIpads, during the trial was the intervention of AFRIpads the only thing they were given, how many, were they free? Did provision stop when the trial was completed, and what is the lag time between completion of the formal trial and this follow up survey? If the trial is over did participants continue receiving them free or have to pay for them? Some studies provide hand wash soap to girls to ensure hygiene (including controls), one assumes this is not so in the study, but worth mentioning. What is the proportion of responders originating from the original trial, and how were they samples, for example is it a 10% random sample of all trial girls? It is not clear if girls surveyed who were not in the trial are different from those who were in the trial (ie is it correct to pool girls in the control or education arm with girls not exposed to the trial participant information sheet, or training at trial onset); would non-trial girls have poorer practice? Was it a classroom survey – was each face to face interview conducted in privacy or were other girls present? (Were girls overheard by their peers?). How long did each interview with a single girl take? Since use of water, soap, washing and drying facilities were important components (particularly drying in or outside), some information about the 8 schools attended by the participants, and their WASH facilities would be relevant. Answers to soap/water were categorical (yes/no), so it is not quite clear how girls interpret the word 'regular' in order to answer yes or no to access? With the washing questions, it is not clear if it refers to washing reusable pads in school, at home or both; how did girls chose to answer, as these responses would vary by place? If washing is at home only, how do girls transport dirty reusables home without stigma? What does (P3-7) mean? Attendance data was recorded for girls participating in the menstruation and the cycle of poverty trial. Are they linked respondents to the survey? If so, how many girls was this? Ethical approval was given for trial study participants, and consent and assent given. Please clarify what ethical approval was granted for non-trial girls who participated in the survey after verbal assent, they had no parental consent, the age range of 10-19y suggests some of these girls could have been quite young. The questions are also quite complex for young girls, was there a difference in ability to answer some questions between the youngest and oldest? Results Participant characteristics – it states 145 were former trial participants of whom 124 received intervention, 87 'pads' – were they AFRIpads? It appears the 21 who were in the trial and were controls, were aggregated with 60 girls who were not in the trial at

	all? Somewhere there needs an explanation on why 60 non-trial girls (who must be different) have been included in this survey. Table 1 - please put the age distribution in the table, due to wide age range; it is unclear what proportion are young girls, or if it is clustered. Is M mean or mode? The authors are keen to highlight the need for prevalence points; with such a small sample size, it would appear relevant to provide (95% confidence limits around the point prevalence). When describing which menstrual products it is still not clear if current availability and use of AFRIpads is trial dependent - as one assumes the trial had finished. Are the AFRIpads now old and worn, were new ones issued? Table 5 was a little unclear, if it includes multivariate analysis, what are the covariates included in the model, and should it not be adjusted (AOR) rather than OR? Age and school controlled for? School appears not to be included as a co-variate, differences in the school environment could account for differences between inadequate and adequate MHM. Discussion Good discussion, and limitations are well defined, although the small sample size needs to be addressed in the limitations, as it prevents any stratification. The effect of age is likely, as younger inexperienced girls are less likely to follow correct practice – were they evenly distributed between the groups? The effect of school externalities as potential confounders requires consideration – how were AFRIpads distributed across schools, are there school measures to ensure they were evenly distributed across the groups, school ideally should be controlled for in multivariate analysis. There is also nothing discussed about diluting the survey by non-trial girls, are they not different in terms of practices compared with trial girls? As MHM of AFRIpads is poor, one argument would be that these trial girls (assuming non-trial girls do not have access to AFRIpads?) were not well trained to use these items correctly, or needed a refresher training since the trial inception. Are the AFRIpads now old and girls do not care for them well, or they chafe more? It appears the 'poor MHM' largely reflects the one question on drying as half of girls hide their pad when drying, while other practices were mostly 'correct' according to the definitions; so the aggregate classed as inadequate MHM is rather overshadowed by this one effect. As the multivariate analysis showed no difference between adequacy of MHM and health markers, it rather weakens the argument, and may be more a consequence of the study sample chosen, and uncontrolled externalities due to the small sample, than a true lack of association. The argument is rather limited without better understanding of the relative contribution of varying school WASH. Was school WASH independently associated with the aggregated 'adequate' MHM measure? One rather obvious discussion point would be presence of water and soap in school, and whether there was a wash room to dry washed reusable materials, all of which are recommended for adequate WASH. References Ref 23 is a report, exchange with the peer-reviewed publication in PLoSMED (Sommer et al 2016). For ref 12, 'association' rather than predict should be used.

REVIEWER	Bethany A. Caruso Emory University, USA
REVIEW RETURNED	24-Jun-2016

GENERAL COMMENTS	Summary This paper provides an important contribution to the literature to date regarding girls' experiences of menstruation with specific attention to the full scope of their menstrual management practices, assessed individually and collectively, and not just the type of absorbent used. The authors assess menstrual hygiene management based on a standardized definition, and such an endeavor could be adapted to other contexts and populations. However, the study as currently presented needs strengthening. The methods need clarification in many places and the limitations of the measure need to be more thoroughly addressed and discussed, as do the outcome measures. As a result, it is difficult to understand some of the findings (specifically the psychosocial measure) and the study could not be duplicated, nor could others be able to adapt the methodology to other areas. Moreover, some indicators are described in a vague way throughout (absenteeism and privacy) and these need to be both described and labeled in a manner so as to not enable any incorrect concussions from readers. The mission of this paper is sound yet time dedicated to strengthening methods and results is needed. Abstract A.1. In the conclusion section, I suggest changing 'prevalence of MHM' to 'prevalence of adequate MHM' as presumably the question is not around whether or not MHM exists, but the quality as operationalized by the definition. A.2. The term 'pooled MHM' is confusing. The reader can get a better sense reading on through the methods, but I suggest alternate language for the abstract. The term could be confused for an aggregate of MHM scores across a population. Intro I.1. While I appreciate the desire to not want to repeat the details of the trial and the interventions, that the manuscript describing the context of this study is not yet out creates a bit of a concern for the reader. I suggest including more information in supplemental material if this one moves forward for publication before the other is published. I.2. p.6, line 10: It is unclear what is meant by '(P3-7)' in the text. I.3. p.6, line 19: Please indicate specifically what the 'small amount of soap' is. (i.e. provide a measure in terms of volume/grams). Methods M.1. Regarding participants: The sample size is confined to participating schools. Please comment on whether or not the study is adequately powered. M.2. Water/Soap: The students were queried about access to water and soap at school from their perspective, which is very useful. Were any school observations made to understand if water and soap were made available. It would be interesting to know if those resources could be visual seen at the school but not be perceived to be regularly accessible by the participants in the study.

	M.3. Drying: The response options for drying include 1 outdoor place (outside), 2 inside places (under bed, inside dorm) and a 'secret place'. It is not entirely clear which, if any, of these is actually an optimal response given that they are all quite vague. From a survey development/measurement perspective, they are not necessarily mutually exclusive. Under the bed or in the dorm could be in secret places, so could an outside space. They could also conceivable have a secret place outside in the sun or their place outside could be under another piece of clothing to conceal their cloth (preventing direct sunlight). You also then indicate that the responses were then back-coded, but it is not clear what responses went into each category given the ambiguity noted. a. Please comment on this question in more depth when discussing the response options, specifically in the MHM measure section. b. Further, given the ambiguity in potential responses, please indicate the limitations of the question as it currently is. c. Please discuss how you may consider refining these responses in future studies for a more robust measure. M.4. p. 8, line 18: "No girls reported…" should be in the results section. M.5. MHM practice questions generally: The survey questions all ask about what girls 'usually' do. Can the authors indicate why a specific time period (i.e. last menstrual period) was not asked about specifically? Asking about 'usual' experiences leaves more room for desirability bias and this should be noted in limitations. M.6. MHM Measure: a. For reader ease, the description of the MHM measure should follow or be combined with the section describing MHM practices. In order to fully assess the MHM measure section, I have needed to flip back and forth between these sections. Given that the primary focus of this paper pertains to the MHM measure, it would be appropriate to have this discussed first in the section, with details then following regarding each component of the measure and the corresponding survey questions. Participant demographics and other measures could be moved to the end. b. Please provide further information as to what is meant by 'pooled'. Was an aggregate score created based on the criteria noted? How was the score generated? M.7. Health. Any citations available to indicate the validity of the questions in this section? M.8. Psychosocial well-being: a. Why is psychosocial well-being not considered health? It is appropriate for this to be under the same heading. b. The presentation of the SDQ should be strengthened. First, flip sentence 'The total score from…" to read that 'Psychosocial well-being was assessed using ….". Next, indicate what this measure is/what its purpose is. Explain if validated and where has it been used in the past/ if it has it been used with populations similar to the one in this study (i.e. among adolescents/in Uganda/etc.). Then give an example or two of the types of questions in the scale, followed by the types of options available for response (Likert info as you have

	it.) Finally, be sure to explain how it is scored (summative?) as well as what the scores mean (i.e. scores below X indicate Y and score above X indicate Z.) M.9. School attendance: As written, this is general attendance. Is there an attendance measure specific to when girls are menstruating? If not, the limitations of this must be noted in the results and discussions and any conclusions drawn must be cautious. Results R.1. p.11, line 9. For the sentence starting with "A total of 124…" and the sentence following: add percentages to numbers. R.2. Table 1: The percentage not providing an answer should be indicated or the N should be adjusted in the table for that subset of questions. R.3. p.11. line 47: The specifics of those who did or did not report on certain items should be noted not in the text but in the table as a footnote. The same strategy should be employed for all tables. R.4. Table 2: For disposal, why are both 'never change at school' and 'board at school' not applicable? Seems that this is an imperative question for girls who board to answer and responses are highly applicable. R.5. p.13., line 5: 'By adding each…" The detail here should be provided in the methods section. R.6. Table 3: a. It needs to be made more clear what 'Usual practices' are. Everything but Afripads? So does that mean this category includes commercial disposable pads and cloth? If so, please justify this grouping. b. Also, it is unclear why results are stratified to have Afripads assessed independently. Is this related to an assessment of the initial trial inputs? If so, please clarify. c. The current format of this table is confusing with the parentheticals and '=' statements. Suggest re-formatting. R.7. Table 4: a. Absenteeism at follow up is indicated as a proposed consequence of poor MHM, however it is not made clear how absenteeism data collected is associated with menstruation as it is just a snapshot in time and follow-up questions about if a girl was menstruating were not carried out. Please justify inclusion of this here or remove. It may be inadvertently interpreted to be absenteeism associated with MHM and this would be inappropriate. b. With the information provided thus far, it is not clear to readers how to interpret the data provided regarding the SDQ. How does one interpret a mean score of SDQ of 19? R.7. p.15., line 36. Regarding attendance and MHM associations, please clarify if this is the actual attendance data or self report of missing during menses. If attendance data, please justify using this data when it may have not been collected when girls were actually

	menstruating. R.8. MHM practices and associations. The justifications for some of the associations tested are not clear. What, for example, is the justification for assessing the association between privacy for washing and standing in class? How would the conditions surrounding a behavior in another pace in time have an impact on a very different behavior in another place in time? When testing associations, there should be some plausibility that one could have an influence on the other. When a large set of associations are tested, it is statistically probable that at least something will come out significant. Without justification for the association tested, it seems to be mining. I would suggest being clearer about the associations tested and the plausibility of why they would be associated in the real world context (not just statistically). Discussion/Conclusion D.1. p. 19, line 17. Drying: Another possible explanation for the unexpected association with drying is that the question and response options were not specific enough. As noted in previous comments, the assumptions made about certain places on behalf of the researchers regarding which locations were safe/clean, may be inaccurate. This should be noted here. D.2. p. 19, line31. Privacy: This should be clarified (here and in table) that privacy is specific to privacy when washing (according to the definition of the researchers). The reader may read this as privacy related to changing, drying, etc. D.3. p. 19, line 50. Attendance: In this paragraph, it is imperative that the researchers make it clear that this measure of attendance being assessed is general attendance and is not specific to attendance during a girls' menstrual period. Readers would have to be savvy to understand this important distinction as it is not made clear enough in the results or the methods. However, the distinction is critical, and the conclusions that can be drawn from any associations made are very very limited because this is not menstruation specific attendance data. I encourage the authors to be more clear and re-think the manner in which this association is discussed in this section. D.4. Limitations. I advise the authors to address the limitation of the individual measures used to make up the pooled measure they presented here. These are noted overall, but it would be useful to be more specific. There are issues with the drying indicator and a few others as noted. In other places in the discussion, the reliability and validity of the measure was questioned. I suggest this be noted here.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Penelope A Phillips-Howard
Institution and Country: Liverpool School of Tropical Medicine, UK
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The study aims to quantify the main components of menstrual hygiene management (MHM),

aggregate respondent's answers to measure if 'adequate' MHM influences some key outcome indicators. Authors note the small sample size, self-reported information, interviewer translation of questions while conducting the face-to-face interviews (with reverse translation for data entry per question), and the cross-sectional nature of the study, all of which affect interpretation of findings. Of note, the survey interviewers could have encountered a desirability effect as they were the same staff that had provided the intervention of AFRIpads in the former trial (it is unclear if they are still distributing, or these are very old pads).
Nevertheless, despite these limitations, the approach toward quantifying information is an important contribution toward this hitherto neglected topic. The results are concerning, particularly the implication that provision of reusable pads did not significantly contribute towards adequate MHM. The implication that AFRIpads do not improve MHM, and that girls who reported adequate MHM did not appear to have significantly improved outcome measures is difficult to explain, and may in part reflect the small sample size, and inability to control for confounding. If the provision of AFRIpads truly did not improve MHM, this has implications for programmes wishing to scale-up provision. The authors suggest disposable pads may be more valuable, as they do not depend on washing and drying, however disposal and cost of such items needs to also be considered in the implications. We note no mention is made of menstrual cups, although they require minimum WASH facilities, are reusable, and cost-effective.
The authors can help readers better understand and interpret the findings, by strengthening parts of the manuscript, with attention to the following suggestions.
Response:
Thank you for taking the time to review this manuscript and provide useful feedback which we believe has greatly strengthened the paper.
We agree with the reviewer as to the limitations of the study, acknowledged in the discussion. We have noted consideration to menstrual cups in the discussion, although fear they would face the same issues with washing and privacy as the reusable pads.
Concerns about sample size have also been added to the 'strengths and limitations' box to further acknowledge this.
We recognise more information about the broader trial needed to be included in this work, particularly since the primary outcomes paper has not been released. We have amended this throughout.
We have addressed specific comments below.

Methods
First line in methods is rather odd, it adds nothing as the specific details of a correct study should be in the methods.
Response:
While we agree that the details of the study should be included in the methods in accordance with the relevant reporting guideline (STROBE in this case), we have included this statement and reference to the check-list in supplementary materials to comply with BMJ Open journal requirements which ask authors to provide a reporting guideline checklist and make clear its use. We have included the reference in text to demonstrate to readers this compliance. Additionally, we feel that in the field of menstrual health studies adherence to reporting guidelines is incredibly poor. We'd like to do whatever possible to provide examples of best practice and direct readers to relevant guidelines.

More detail is needed about the trial. It would be useful to know if girls provided with AFRIpads were truly randomised in the trial, it states the trial was quasi-randomised, are the girls in the control arm different from those not receiving AFRIpads? Was school the unit of randomisation?
Did the authors perform sample size calculations for which the small sample was chosen, particularly to allow stratification by age and school?
The study population is 205 girls, I can see no reason for mentioning the non-menstruating girls who were non-eligible.
It was not clear to the reviewer if provision of AFRIpads was ongoing, and if the trial was completed. It

is difficult to define if it is part of the trial, or is a stand-alone cross sectional survey, after closure. It is relevant to know how long girls were provided AFRIpads, during the trial was the intervention of AFRIpads the only thing they were given, how many, were they free? Did provision stop when the trial was completed, and what is the lag time between completion of the formal trial and this follow up survey? If the trial is over did participants continue receiving them free or have to pay for them? Some studies provide hand wash soap to girls to ensure hygiene (including controls), one assumes this is not so in the study, but worth mentioning.
What is the proportion of responders originating from the original trial, and how were they samples, for example is it a 10% random sample of all trial girls? It is not clear if girls surveyed who were not in the trial are different from those who were in the trial (ie is it correct to pool girls in the control or education arm with girls not exposed to the trial participant information sheet, or training at trial onset); would non-trial girls have poorer practice?
Response:
Thank you for providing this feedback. While these details are provided in the primary trial paper (which we anticipated to be released before this follow-up), we realise that more detail is needed for this stand-alone paper. We have increased detail in this paper, although have tried to avoid excessive repetition.
This paper uses data from the final survey of the trial. It was not a separate study and an independent sample size calculation was not undertaken. We have made explicit the relationship of the 'non-trial' girls in the summary in the Introduction, and expanded some parts of the Methods to make this really clear.
As added to the introduction:
Girls who transferred into the study schools during the trial were included in intervention delivery (if attending school at those times) and follow-up surveys as not to identify or stigmatise girls in the trial or discriminate against those transferring in after baseline from receiving resources. While girls transferring into the school could not be included in primary trial intention-to-treat analyses, their responses were incorporated into the final survey data set for secondary analysis. This maximised cross-sectional sample size in the survey data set.
We would note to the reviewer that many of these girls transferred to the trial schools as early as during the baseline attendance data collection. Understandably we used an intention-to-treat design for the trial and these girls transferring in were not included in the 'trial'. But many had been at the school almost as long as the 'trial' girls and received the interventions.
To improve clarity, throughout the paper we have revised to refer to this as a secondary analysis of the final survey data. We hope this will avoid reader confusion, particularly given delays in the primary trial paper release.

We have included additional information about the trial in the present study section of the Introduction including details of the AFRIpad packs provided and dates. As stated in the Introduction, the trial was undertaken between January 2012 and December 2014.
AFRIpads were provided in October 2012 and March 2014, girls were provided with a pack of 6 reusable pads.
The final survey was conducted in October/November 2014 so the latest pads were about 7/8 months old. We have also noted this in the discussion.
We have included the 435 girls figure for consistency with the primary trial paper so it is clear where the papers connect. We realise this is not available yet but feel it is important to leave this here, the abstract uses the 205 figure to be clear about the n of this study.

Was it a classroom survey – was each face to face interview conducted in privacy or were other girls present? (Were girls overheard by their peers?). How long did each interview with a single girl take?
Response:
We have added this detail to the methods. Interviews were around 30-40 minutes and conducted individually in a private place.

Since use of water, soap, washing and drying facilities were important components (particularly drying in or outside), some information about the 8 schools attended by the participants, and their WASH facilities would be relevant.
Answers to soap/water were categorical (yes/no), so it is not quite clear how girls interpret the word 'regular' in order to answer yes or no to access?
Response:
'Regular' access to water should be improved in future surveys, this was interpreted by the interviewers as that girls rarely did not have access to water. We have discussed measurement issues and recommendations for future work in the discussion.
No girls reported washing or drying their sanitary pads at school (Table 2). Given the emphasis in the field on WASH in schools we have addressed this in the discussion. However, home WASH would have been much more important for this sample since this is where they washed and dried their absorbents.

With the washing questions, it is not clear if it refers to washing reusable pads in school, at home or both; how did girls chose to answer, as these responses would vary by place? If washing is at home only, how do girls transport dirty reusable home without stigma?
Response:
Girls in the sample all report taking their absorbents home to wash them. This is reflected in answers to questions in Table 2, and also in our site visits and qualitative interviews with the girls. We have highlighted this in text, particularly in the discussion. We've also briefly addressed WASH comparability in schools in the introduction. Transporting absorbents is a really important point to note and we've included this in revised discussion.

What does (P3-7) mean?
Response:
This was the primary grades included in the primary school (standard for Uganda), we have clarified this in text. Information presented is consistent with the breakdown across grades in Table 1 and the primary trial paper.

Attendance data was recorded for girls participating in the menstruation and the cycle of poverty trial. Are they linked respondents to the survey? If so, how many girls was this?
Response:
As above we've added clarity around this in introduction and methods.
As stated in results 145 of the 205 girls had been in the study from baseline (and thus had attendance data). We have also added this figure to the methods where methods of attendance data collection were reported. We have noted that attendance data was linked to survey responses through participant IDs. As noted above we've also improved clarity that this study was conducted on the survey data from the trial, not a separate study/survey.

Ethical approval was given for trial study participants, and consent and assent given. Please clarify what ethical approval was granted for non-trial girls who participated in the survey after verbal assent, they had no parental consent, the age range of 10-19y suggests some of these girls could have been quite young.
The questions are also quite complex for young girls, was there a difference in ability to answer some questions between the youngest and oldest?
Response:
We have added the age distribution to Table 1. There was 1 girl who reported being 10 and 1 who reported being 11. As questionnaires were verbally administered in the local language research assistants were trained to assist girls in understanding the questions as well as informed consent. Ethical approval for the trial included all girls in the trial and coming to the schools. Further as part of

the broader study ethical approval was provided for household surveys, community group meetings. Throughout the trial there were stakeholder meetings with parents from the community. School principals all provided consent for the study. The partner NGO was well known to schools throughout the area and has excellent relationships with the community leaders.
Ethical approval was granted to survey all girls in the study schools, which extended to non-trial girls as more girls came to the schools. It was considered unethical to exclude girls transferring into the schools, particularly for pad provision.

Results
Participant characteristics – it states 145 were former trial participants of whom 124 received intervention, 87 'pads' – were they AFRIpads? It appears the 21 who were in the trial and were controls, were aggregated with 60 girls who were not in the trial at all? Somewhere there needs an explanation on why 60 non-trial girls (who must be different) have been included in this survey.
Response:
This survey was the follow-up survey from the trial. It was not a separate survey and this has been clarified in revisions. The 60 girls who were also included in this study had transferred to the study schools during the course of the trial. Of these, 28 had received an intervention (26 received AFRIpads). We agree this needed to be clearer, particularly in absence of the availability of the trial paper. As noted above, this has been added to introduction, methods and discussion.
As we discuss at length in the trial paper, there was incredibly high transfer rates out of, and into, the study schools. Girls included in this cross-sectional follow up, who were not in the trial at baseline, show no differences to those who were in the schools at baseline.

Table 1 - please put the age distribution in the table, due to wide age range; it is unclear what proportion are young girls, or if it is clustered. Is M mean or mode?
Response:
We have added the age distribution to Table 1, as seen there 87% of the sample were between 13 and 15. This is the mean and we have added the full word in text.

The authors are keen to highlight the need for prevalence points; with such a small sample size, it would appear relevant to provide (95% confidence limits around the point prevalence).
Response:
We have added confidence intervals to reported proportions of those with adequate and inadequate MHM. Although would stress the study only claims that this is the prevalence of menstrual hygiene in the study sample. The findings are important, as is the theoretical contribution of the paper in attempting to provide a single prevalence estimate of the MHM construct.

When describing which menstrual products it is still not clear if current availability and use of AFRIpads is trial dependent - as one assumes the trial had finished. Are the AFRIpads now old and worn, were new ones issued?
Response:
We have included more references to the dates in the trial. Further we have improved clarity in text that only those in conditions receiving the AFRIpads would have had access to them. However, we note that the impact of AFRIpad provision on menstrual hygiene is not the primary objective or focus of the study. As emphasised in the paper, this work serves primarily to assess the prevalence of MHM in a way that is consistent with the content definition and provide a worked example of the quantitative assessment of MHM and its association with hypothesised consequences. In this way the theoretical contribution of the paper extends beyond the results presented for this smaller sample size.
From the last provision of AFRIpads until the time of the survey, the time difference was 7-8 months. AFRIpads should last for 12 months and this would have given girls sufficient time to adjust to using the AFRIpads. After the final surveys were conducted, all schools (including control schools) received the pads.

Table 5 was a little unclear, if it includes multivariate analysis, what are the covariates included in the model, and should it not be adjusted (AOR) rather than OR? Age and school controlled for? School appears not to be included as a co-variate, differences in the school environment could account for differences between inadequate and adequate MHM.
Response:
We have added a footnote to this table indicated that the adjusted ORs are where all 4 measured aspects of MHM were included in the model. We have revised the analyses section in the methods to be more clear about this.
Schools were all comparable in their WASH and other characteristics. As discussed above, girls did not wash or dry their absorbents at school. While time since menarche, rather than age, might be a useful covariate to consider in girls' improved menstrual hygiene, age may not also be associated with health issues or the experience of shame. There is very limited existing evidence to base covariates on. If possible we believe parents' education and poverty indexes would be the most likely confounds between MHM and the outcomes that we've tested.
Given the small sample size and the fact that AFRIpad provision was dependent on school, models could not accommodate adding these covariates and we have addressed the need for confounder adjustment in future studies in the revised discussion.

Discussion
Good discussion, and limitations are well defined, although the small sample size needs to be addressed in the limitations, as it prevents any stratification. The effect of age is likely, as younger inexperienced girls are less likely to follow correct practice – were they evenly distributed between the groups? The effect of school externalities as potential confounders requires consideration – how were AFRIpads distributed across schools, are there school measures to ensure they were evenly distributed across the groups, school ideally should be controlled for in multivariate analysis.
Response:
Thank you. We have further stressed sample size limitations in the discussion and added this to the 'strengths and weaknesses' box with the abstract to highlight it further.
AFRIpads were distributed in the two conditions (so 4 schools) quasi-randomised to receive the pads.

There is also nothing discussed about diluting the survey by non-trial girls, are they not different in terms of practices compared with trial girls?
Response:
We have addressed this in comments above and added clarity to introduction and methods, we've also noted this in the discussion.

As MHM of AFRIpads is poor, one argument would be that these trial girls (assuming non-trial girls do not have access to AFRIpads?) were not well trained to use these items correctly, or needed a refresher training since the trial inception. Are the AFRIpads now old and girls do not care for them well, or they chafe more?
Response:
Again we would stress that the comparison with those who were using AFRIpads was not the primary objective of the study. It is an important finding, however, and issues around AFRIpad use training are important. We have added this to the discussion – thank you.
We would suspect though that other constraints facing girls such as the inadequacies of latrines and access to private water and drying options at home (rather than the education provided with the pads) might be more likely to be responsible for this effect.

It appears the 'poor MHM' largely reflects the one question on drying as half of girls hide their pad when drying, while other practices were mostly 'correct' according to the definitions; so the aggregate classed as inadequate MHM is rather overshadowed by this one effect. As the multivariate analysis

showed no difference between adequacy of MHM and health markers, it rather weakens the argument, and may be more a consequence of the study sample chosen, and uncontrolled externalities due to the small sample, than a true lack of association. The argument is rather limited without better understanding of the relative contribution of varying school WASH. Was school WASH independently associated with the aggregated 'adequate' MHM measure? One rather obvious discussion point would be presence of water and soap in school, and whether there was a wash room to dry washed reusable materials, all of which are recommended for adequate WASH.
Response:
While drying does contribute significantly to 'poor MHM' all aspects played a role. Further, if the field defines MHM as it has, why should this aspect be any less important when trying to provide a prevalence estimate of poor MHM for this population? We would note the incredibly high correlation between those who report inadequate drying and those who report often or sometimes wearing pads damp as an indication that this is a real reflection of behaviour rather than a facet of the question. Given few studies have tested associations between aspects of MHM and health markers, it would be difficult to say that this weakens the argument here as no association may exist. Indeed, the Das et al study found reusable pads to be associated with increased confirmed urogenital infection. We do see the small sample size and very poor MHM amongst this sample as an issue, but it is still important to present the results of the comparison for this group.

We have noted the comparability of WASH in the schools across all 8 schools in revisions. Since girls did not wash or dry their pads at school this would not have had implications for their MHM. Schools had almost identical latrines so it is also unlikely this impacted on changing frequency.
While interventions improving school based WASH to the point that girls handled their MHM in schools might result in drastic improvements to MHM (we've never seen a trial or quantitative cross-sectional study on this), this was not the focus of the present study or the broader trial. We have noted school WASH issues in the discussion.

References
Ref 23 is a report, exchange with the peer-reviewed publication in PLoSMED (Sommer et al 2016).
For ref 12, 'association' rather than predict should be used.
Response:
This has been amended.

Reviewer: 2
Reviewer Name: Bethany A. Caruso
Institution and Country: Emory University, USA
Please state any competing interests or state 'None declared': None Declared

Please leave your comments for the authors below
Summary
This paper provides an important contribution to the literature to date regarding girls' experiences of menstruation with specific attention to the full scope of their menstrual management practices, assessed individually and collectively, and not just the type of absorbent used. The authors assess menstrual hygiene management based on a standardized definition, and such an endeavor could be adapted to other contexts and populations. However, the study as currently presented needs strengthening. The methods need clarification in many places and the limitations of the measure need to be more thoroughly addressed and discussed, as do the outcome measures. As a result, it is difficult to understand some of the findings (specifically the psychosocial measure) and the study could not be duplicated, nor could others be able to adapt the methodology to other areas. Moreover, some indicators are described in a vague way throughout (absenteeism and privacy) and these need to be both described and labeled in a manner so as to not enable any incorrect concussions from

readers. The mission of this paper is sound yet time dedicated to strengthening methods and results is needed.
Response:
Thank you for taking the time to review the manuscript and provide useful feedback. We have incorporated changes – see below. We have added notes that privacy in the tables refers to privacy for washing to improve clarity around this.

Abstract
A.1. In the conclusion section, I suggest changing 'prevalence of MHM' to 'prevalence of adequate MHM' as presumably the question is not around whether or not MHM exists, but the quality as operationalized by the definition.
A.2. The term 'pooled MHM' is confusing. The reader can get a better sense reading on through the methods, but I suggest alternate language for the abstract. The term could be confused for an aggregate of MHM scores across a population.
Response:
We have included these changes in revisions to the abstract, thank you.

Intro
I.1. While I appreciate the desire to not want to repeat the details of the trial and the interventions, that the manuscript describing the context of this study is not yet out creates a bit of a concern for the reader. I suggest including more information in supplemental material if this one moves forward for publication before the other is published.
Response:
Thank you, we agree. Unfortunately, the review process for the main trial paper is taking much longer than anticipated (despite being submitted well before this manuscript). We are hesitant to repeat too much detail but have added more information about the trial in Introduction and Methods. We believe this has improved clarity around the trial methods for understanding this paper in context of the broader work.

I.2. p.6, line 10: It is unclear what is meant by '(P3-7)' in the text.
Response:
This was the primary grades included in the primary school (standard for Uganda), we have clarified this in text. Information presented is consistent with the breakdown across grades in Table 1 and the primary trial paper.

I.3. p.6, line 19: Please indicate specifically what the 'small amount of soap' is. (i.e. provide a measure in terms of volume/grams).
Response:
We have added that this was one sachet (45 grams).

Methods
M.1. Regarding participants: The sample size is confined to participating schools. Please comment on whether or not the study is adequately powered.
Response:
The present study is a secondary analysis of the final survey from the primary trial. A-priori power analysis was conducted for the trial but this study used the relevant data available, thus no a-priori power analysis was conducted.
We have added comment on power issues in the strengths and limitations section of the discussion. This study was the first to assess relationships between many aspects of MHM and outcomes, and to provide a combined measure of MHM. It would have been very difficult to conduct meaningful power analysis for this study as there was limited previous literature on which to anticipate the proportion of

girls meeting criteria at each step, or overall MHM. Secondly, there are very limited studies on which we could have generated anticipated effect sizes of the impact of aspects of MHM on outcomes. We added this note to the discussion.

M.2. Water/Soap: The students were queried about access to water and soap at school from their perspective, which is very useful. Were any school observations made to understand if water and soap were made available. It would be interesting to know if those resources could be visual seen at the school but not be perceived to be regularly accessible by the participants in the study.
Response:
We undertook site visits at each of the schools. No schools provided soap. Some schools were closer to water sources than others but water was not provided in a way that would have been helpful for MHM. No girls washed their absorbents at schools (Table 2). We've added extended discussion around WASH in schools to the Discussion of the paper.

M.3. Drying: The response options for drying include 1 outdoor place (outside), 2 inside places (under bed, inside dorm) and a 'secret place'. It is not entirely clear which, if any, of these is actually an optimal response given that they are all quite vague. From a survey development/measurement perspective, they are not necessarily mutually exclusive. Under the bed or in the dorm could be in secret places, so could an outside space. They could also conceivable have a secret place outside in the sun or their place outside could be under another piece of clothing to conceal their cloth (preventing direct sunlight). You also then indicate that the responses were then back-coded, but it is not clear what responses went into each category given the ambiguity noted.
a. Please comment on this question in more depth when discussing the response options, specifically in the MHM measure section.
b. Further, given the ambiguity in potential responses, please indicate the limitations of the question as it currently is.
c. Please discuss how you may consider refining these responses in future studies for a more robust measure.
Response:
We completely agree with the reviewer that the question was not ideal for the purposes of this study and should be improved in future. We have incorporated more discussion of questions in the discussion section of the paper. We do note that all girls who reported often or sometimes wearing absorbents damp dried them hidden, which adds some validity to this item.
Note we included the survey measure (in full) in the supplementary materials (Supp materials 2) to this paper to aid in full transparency.

M.4. p. 8, line 18: "No girls reported…" should be in the results section.
Response:
This has been moved in revisions.

M.5. MHM practice questions generally: The survey questions all ask about what girls 'usually' do. Can the authors indicate why a specific time period (i.e. last menstrual period) was not asked about specifically? Asking about 'usual' experiences leaves more room for desirability bias and this should be noted in limitations.
Response:
Again we agree and think work needs to be done around this, this is also addressed in expanded discussion.

M.6. MHM Measure:
a. For reader ease, the description of the MHM measure should follow or be combined with the section describing MHM practices. In order to fully assess the MHM measure section, I have needed to flip back and forth between these sections. Given that the primary focus of this paper pertains to

the MHM measure, it would be appropriate to have this discussed first in the section, with details then following regarding each component of the measure and the corresponding survey questions. Participant demographics and other measures could be moved to the end.
Response:
We have reordered this section as suggested. We felt it still valuable to have the description of items separate from the criteria of the MHM measure to avoid confusion, particularly given two sets of criteria are provided (for MHM minimal and relaxed criteria), but have moved the description of the items higher in the methods section with the description of MHM criteria following directly afterwards.

b. Please provide further information as to what is meant by 'pooled'. Was an aggregate score created based on the criteria noted? How was the score generated?
Response:
This is an aggregated/pooled amount – this is displayed in Table 3. We've added some more signposting in this section of the methods highlighting links to the aspects of MHM discussed in the background section of the paper. The score was generated by applying each criterion for MHM sequentially.

M.7. Health. Any citations available to indicate the validity of the questions in this section?
Response:
We have discussed limitations of self-reported health symptoms in the discussion of the paper. Questions were developed based on the Ghana pilot and qualitative work as stated in the survey design section. Very little has been done to validate questions in this area. We did certainly consult the Sumpter & Torondel review, but there simply haven't been rigorous validity studies for (really any) MHM measures.

M.8. Psychosocial well-being:
a. Why is psychosocial well-being not considered health? It is appropriate for this to be under the same heading.
Response:
We agree that this could be considered health but think that separating these items is useful. It is also useful to break up the large table.

b. The presentation of the SDQ should be strengthened. First, flip sentence 'The total score from…" to read that 'Psychosocial well-being was assessed using ….". Next, indicate what this measure is/what its purpose is. Explain if validated and where has it been used in the past/ if it has it been used with populations similar to the one in this study (i.e. among adolescents/in Uganda/etc.). Then give an example or two of the types of questions in the scale, followed by the types of options available for response (Likert info as you have it.) Finally, be sure to explain how it is scored (summative?) as well as what the scores mean (i.e. scores below X indicate Y and score above X indicate Z.)
Response:
Thank you for this comment. The primary trial paper includes a far more detailed description of the SDQ use, as well as results and the implications of these. Our intention here was not to duplicate this information. However, it is clear from these comments that more is needed for this paper. We have added example items for each subscale, note that scores are summed, and reference to the free online website for the SDQ which includes cut offs for many countries, different version of the SDQ, and a wealth of information about the tool. We have added an extra comment in the discussion about the high SDQ scores found in this sample.

M.9. School attendance: As written, this is general attendance. Is there an attendance measure specific to when girls are menstruating? If not, the limitations of this must be noted in the results and discussions and any conclusions drawn must be cautious.
Response:

This is general attendance and we have clarified this.

Results
R.1. p.11, line 9. For the sentence starting with “A total of 124…” and the sentence following: add percentages to numbers.
Response:
This has been added in revisions.

R.2. Table 1: The percentage not providing an answer should be indicated or the N should be adjusted in the table for that subset of questions.
Response:
This has already been done throughout the paper. Where the n differs from that stated at the top of the table it has been provided in brackets or there is a specific ‘no answer’ figure in italics. Where the N differs not due to non-response, but due to the question being filtered on a previous question this has been noted in footnotes (e.g., see Table 2).

R.3. p.11. line 47: The specifics of those who did or did not report on certain items should be noted not in the text but in the table as a footnote. The same strategy should be employed for all tables.
Response:
For everything except the 4 girls who did not report on absorbents used this has been indicated in the table with a footnote. However, for these 4 girls they are not included in Table 2, 3 or 4, because there wasn’t sufficient information about absorbent used which is then important for all follow-up questions. It would be laborious to add this footnote to all tables, which already include a lot of other footnotes about who did or did not answer questions based on survey filters. This is reflected in the n=201 at the top of table 2.

R.4. Table 2: For disposal, why are both ‘never change at school’ and ‘board at school’ not applicable? Seems that this is an imperative question for girls who board to answer and responses are highly applicable.
Response:
This was a poor label choice. The question was about disposal, and for these 3 girls they had volunteered responses that were typed into the survey simply stating that they boarded at school (n=1) or never changed at school (n=2). Since this wasn’t a response option we’ve relabelled this category as voluntary responses. Given the small number, and that we don’t have any more information there’s little more we can do with these 3 responses.

R.5. p.13., line 5: ‘By adding each…” The detail here should be provided in the methods section.
Response:
We’ve added a sentence summarising this in methods. Although we realise it isn’t standard practice to have some of this detail in the results rather than methods, we think it is really useful to have this information here above the table to avoid readers needing to flip back to the methods to interpret the table (also we think many readers frequently skip the methods section). We think it is very useful to have the information right there next to the table, but have added a version of this sentence to methods to address this comment.

R.6. Table 3:
a. It needs to be made more clear what ‘Usual practices’ are. Everything but Afripads? So does that mean this category includes commercial disposable pads and cloth? If so, please justify this grouping.
b. Also, it is unclear why results are stratified to have Afripads assessed independently. Is this related to an assessment of the initial trial inputs? If so, please clarify.
c. The current format of this table is confusing with the parentheticals and ‘=’ statements. Suggest re-formatting.

Response:
Usual practice does include everything except AFRIpads. In this sample this includes a very small number of girls using disposable pads. Essentially it is 'treatment as usual' ie. Everyone who was not using AFRIpads provided as part of the trial.
We realise it is a bit of an unconventional table, but think it is important to have the reminder of each step right there in the table, not just above or in methods. We've changed the '=' to 'criteria' to aid interpretation.
AFRIpads need to be assessed separately as these were provided as part of the Menstruation and the cycle of poverty trial. Thus they were not what girls would have been doing in absence of the trial. It is important to provide the two groups separately to acknowledge this.

R.7. Table 4:
a. Absenteeism at follow up is indicated as a proposed consequence of poor MHM, however it is not made clear how absenteeism data collected is associated with menstruation as it is just a snapshot in time and follow-up questions about if a girl was menstruating were not carried out. Please justify inclusion of this here or remove. It may be inadvertently interpreted to be absenteeism associated with MHM and this would be inappropriate.
Response:
This is general attendance, as is the primary outcome of the Menstruation and the cycle of poverty trial. This is also fairly common practice in studies of menstrual hygiene interventions (e.g., Montgomery et al., 2012; Wilson et al., 2014, even the Oster & Thornton 2011 paper uses general attendance, although did do an assessment using menstrual calendars). We've not seen a study of menstrual-specific attendance.
We have clarified this in methods that this is general attendance & absenteeism which will include both menstruating and non-menstruating days.

b. With the information provided thus far, it is not clear to readers how to interpret the data provided regarding the SDQ. How does one interpret a mean score of SDQ of 19?
Response:
As responded to above.

R.7. p.15., line 36. Regarding attendance and MHM associations, please clarify if this is the actual attendance data or self report of missing during menses. If attendance data, please justify using this data when it may have not been collected when girls were actually menstruating.
Response:
As reported in the methods this is attendance recorded as part of the Menstruation and the cycle of poverty trial – this was collected by trained research assistants from the partner NGO who surveyed 1 week per term for 3 terms in 2014. It was not self-reported. As above recording all attendance days is common practice in MHM intervention studies. It would have been very difficult and intensive to ask girls in the trial (1124 in the bigger trial) to record their menstrual days for the 2 year trial period. General attendance will include days when girls are menstruating. Differences may be diluted with non-menstrual days meaning larger samples are needed, but does not threaten the validity of the measure when comparing attendance based on a menstruation or menstrual hygiene related variable.

R.8. MHM practices and associations. The justifications for some of the associations tested are not clear. What, for example, is the justification for assessing the association between privacy for washing and standing in class? How would the conditions surrounding a behavior in another pace in time have an impact on a very different behavior in another place in time? When testing associations, there should be some plausibility that one could have an influence on the other. When a large set of associations are tested, it is statistically probable that at least something will come out significant. Without justification for the association tested, it seems to be mining. I would suggest being clearer about the associations tested and the plausibility of why they would be associated in the real world

context (not just statistically).
Response:
We believe that all tested associations have relevance to the current working theory of MHM. For example, having privacy to wash absorbents may mean girls are able to do this more thoroughly. This would contribute to hygiene, the performance of their absorbents over time, and perhaps their general anxiety around menstruation. We would assert that testing these associations is consistent with statements in the literature that MHM is linked with 'health, education and psychosocial outcomes'. If all aspects of MHM are not actually linked to all consequences, then perhaps the definition of the concept would need to be addressed. We believe all aspects considered could be associated with consequences assessed.

Discussion/Conclusion
D.1. p. 19, line 17. Drying: Another possible explanation for the unexpected association with drying is that the question and response options were not specific enough. As noted in previous comments, the assumptions made about certain places on behalf of the researchers regarding which locations were safe/clean, may be inaccurate. This should be noted here.
Response:
We have added this, as well as discussion around the items used later in the discussion in revisions.

D.2. p. 19, line31. Privacy: This should be clarified (here and in table) that privacy is specific to privacy when washing (according to the definition of the researchers). The reader may read this as privacy related to changing, drying, etc.
Response:
We have amended this in tables and text.

D.3. p. 19, line 50. Attendance: In this paragraph, it is imperative that the researchers make it clear that this measure of attendance being assessed is general attendance and is not specific to attendance during a girls' menstrual period. Readers would have to be savvy to understand this important distinction as it is not made clear enough in the results or the methods. However, the distinction is critical, and the conclusions that can be drawn from any associations made are very very limited because this is not menstruation specific attendance data. I encourage the authors to be more clear and re-think the manner in which this association is discussed in this section.
Response:
We have attended to this in text and made this explicit in the methods. However, again would note that this has been standard practice in MHM studies and don't think that it would be surprising to readers that this was general attendance.

D.4. Limitations. I advise the authors to address the limitation of the individual measures used to make up the pooled measure they presented here. These are noted overall, but it would be useful to be more specific. There are issues with the drying indicator and a few others as noted. In other places in the discussion, the reliability and validity of the measure was questioned. I suggest this be noted here.
Response:
We have expanded discussion of measurement issues in the discussion.

VERSION 2 – REVIEW

REVIEWER	PA Phillips-Howard Liverpool School of Tropical Medicine
REVIEW RETURNED	19-Aug-2016

GENERAL COMMENTS	The authors have attended to most of the reviewer comments,

	however, there are a number of clarifications and edits now raised after the authors report this study population is part of the trial, and not an independent study. Main comments There is no publication for the trial reference (Ref 24), which the authors note is taking time, but it is frequently used to cover elements of this study. The authors state this study is not separate from the trial, and data are predicated on the trial methods with the design aiming to assess relationships with outcomes compared across arms (and the title reports 'impact'). It is thus important that more detail on the trial methods is provided, it remains insufficient for a stand-alone paper comparing outcomes. It does not matter that there is duplication. Examples of methods requiring more detail include the sample size calculations, inclusion/exclusion criteria for recruitment of the girls and recruitment of the schools, baseline measures of participants, and measures on intervention rigor. A copy of the trial protocol is also required to accompany the review of this submission, and as an online supplement, as results are comparative between arms. I recommend removing the late entrant (transfers). If the authors are to include the 60 transferred girls, not enrolled at baseline they need to better characterize the girls to show they are the same – it looks as if they have added them just to boost numbers, without indicating if these girls were 'treated' exactly the same regarding interaction with research and implementing staff, training, provision of interventions etc. This is discussed further below, particularly regarding intervention allocation when conducting the trial / study ITT analysis. Data on actual menstruation (both base line and end line measures) needs clarification, especially with reported inclusion of very young girls (10-11y olds) at endline. Menstruation must have been an eligibility criteria to enter the trial / study. Girls who were not menstruating at baseline should be excluded or only included once they start menstruating. The strengths and weaknesses section on the title page needs amending to take into account the weaknesses in this trial / study, noted below. The current first and second (and perhaps also third) can be combined, providing room for reporting the weaknesses. The main weakness are detailed below, including inclusion of non-trial girls due to small trial sample size and intervention allocation, poor school attendance measures, information collected on menarche and menstruation (see comments on result, Table 1). Other comments Page 2: the conclusions on AfriPad may not be accurate – more detail is needed on whether girls received the AfriPads, whether they were trained adequately, whether the additional 60 girls received intervention exactly the same as the 145. Page 7 (Survey design): The design is an end line cross-sectional

	survey nested within a trial. why are the controls missing from this trial / study? Were late entrants (transfers) included in the design? Page 7 (Sample size calculation): STROBE wrongly checks this as presented. With due respect to the authors response, sample size calculations are needed. They can be the sample size calculations in the trial protocol, which aim to measure the outcomes reported in this paper. It is clear this is a limitation (but if correctly presented does not preclude publication) and should be included as a weakness of the study in the strengths and weaknesses section in the title page. Page 7 (Intervention methods): The new edits bring up some questions on trial interventions which are relevant to participant responses to MHM. A section in methods on how the interventions are provided is required. Attention to how it was provided to the 29% of sample who are transfers is also needed (including the 11 in Tale 1 who were at a different school the previous year). While the authors state the AfriPads are not the main outcome, a major finding in the abstract and results relate to comparison between arms, showing no difference in MHM between girls in the AfriPads and control arms. It thus remains important to clarify whether AfriPad provision was adequate for these analyses, both in terms of optimal provision, training, and the proportion of girls within this arm receiving the pads. One AfriPad pack containing 2 reusable pads was given in October 2012, and then nothing until March 2014 – this means that girls had 2 reusable pads to last for 18 months. This seems very meagre, other reusable pads are stated to last 4-6 months before disintegrating. Can the authors confirm that AfriPads have a 'shelf-life' of 18 months and remain fully effective for MHM? If not this needs to be discussed in the limitations and discussion, as it would seem obvious that girls will not continue to use old, poor quality materials, resulting in poor MHM (i.e. an inadequate intervention adds to poor MHM, rather than resolving it). This should also be added as a weakness in the strengths and weaknesses section in the title page. please confirm menstruation is an eligibility criteria. The new edits state that girls were provided the AfriPads in the intervention group on recruitment, and received training to use them. Was this conducted at one time-point for all girls? It is much less clear if girls transferring into the school received the intervention with the same training activities. The authors note that the trial analysis of outcomes between arms is intention to treat analysis – which is usually based on provision at enrolment after screening eligibility. If new entrants come later to a cluster, are these girls classed in the trial as eligible for enrolment through rolling enrolment? It is assumed they are included to increase sample size. If they arrive (transferred) from another study school they usually hold allocation for their previous school. This trial appears to allocate the new entrants in the intervention of the new school. The authors need to confirm they are similar in terms of allocation and training; otherwise they would dilute the prevalence of use at the final survey, and reduce the impact of the intervention. The number who transferred after the baseline is not a small portion, with 60 of the 205 girls recorded as late entrants (29%). While the limitations section says 'many' of these were given the intervention; the number and portion is not given – authors need to specify how many received the

	AfriPads and the correct training. Further, how do authors know the late entrants had the same characteristics at baseline – what measures do they have to confirm this? Similarly, it is unclear whether any of the 60 girls recruited later who entered education arms received the same puberty education that the NGO provided as standard to all girls after baseline observations in all girls. Can the authors verify this? If not, the value of including these additional girls is doubtful. Unless there is strong evidence of similar allocation for late entrants, it would be better for them to be excluded or be classed as controls to prevent them diluting the intervention effect. While more detail on this is required, these issues need to be discussed in the 'limitations' section, and possibly as a weakness in the strengths and weaknesses section in the title page. Was the second provision at 18 months at the finish of the trial? Was this to controls or just the pad users – was there an improvement in use once they received the new pads? Page 9 (Survey measures): Measures of baseline characteristics required. Page 10 (Survey measures; health): there are questions on dealing with periods but authors need to include the (biological) menstrual questions i.e. have you ever menstruated (Y/N), when did your menstruation start (date), is your menstruation light or heavy etc. this needs to be reported at baseline and end line. My worry is, that girls not menstruating may slip through and answer 'No' for questions on dealing with their period (i.e. did you use AfriPads – No). Authors state all the 205 girls reached menarche – was this at baseline? Can the authors confirm if the eligibility criteria was all menstruating girls at baseline? Is this in the protocol? Ages are confusing for endline measures – see below on results comments. If girls who were not menstruating at baseline are included this needs to be clearly stated. Surely 10-13 year olds at endline would not be menstruating at trial / study start, and likely a high proportion of these would only just have begun menstruation toward the end of the study; such girls would have little experience of MHM practicalities, so would be different and dilute the outcomes. At the moment only follow-up data is mentioned; age at menarche at the beginning of the study would be useful as it would allow exploration of menstrual experience which many explain differences in ability to deal with MHM and also use of AfriPads. (It appears this was collected from trial girls at baseline - as age at menarche is noted at the beginning of results – was it collected also from the additional 60, so all 205 girls?). Maybe girls who had been menstruating for a number of years were better able to deal with AfriPads compared with girls who started menstruating just before the trial began? Other characteristics measured for end line are very thoroughly reported so it is strange to miss the menstrual questions which are at the heart of the trial study. Page 11 (Survey measures; school attendance): elsewhere it states school registry was used for assessing absence at baseline for the trial. Here the authors switch and say for the present study (but the authors have said it is not a separate study, but same as the trial). Baseline attendance using registers would not have captured the 60 transferred girls added to the 145 trial participants. The authors say

	they have responded to Reviewer 1 and 2 regarding the attendance spot-check – stating that it is common to use general attendance, not menstrual specific. They have not dealt with the more important concern of switching methods of measuring attendance, or the ability of spot-checks rather than continuous surveillance to adequately document this important outcome. In the discussion they state it is not logistically feasible to do continuous attendance monitoring. Use and reliability of spot-check needs to be reported as a weakness in the strengths and weaknesses section after the title page. Non-collection of absence in menstruation is also a weakness. Page 12 (Analysis): In the authors' response to R2, they state SDQ analysis was conducted in detail in the trial paper – is this the same data? The full description is needed as this is a standalone paper. Is all analysis like the trial, using intent to treat – is that in the protocol? Inclusion of transferred girls pooled with the trial girls seems still to be only done to increase sample size. Have the authors conducted analysis without the transferred girls? a supplementary table is needed to show these girls 'are the same', which the authors state. This should include a comparison of baseline and of end line characteristics. Page 13 (Ethics): The authors are requested to provide the protocols. Was the process of consent from the parent and assent from each girl the same procedure for transferred girls – it states girls at the start of the study gave verbal consent – is that different from the written consent they had already given? Again would this miss transferred girls? Pages 14/5 (Results; participant characteristics, Table 1): this appears to be end line participant characteristics. The first sentence is two sentences – the first half states 205 girls were menstruating (is this at the beginning or at the end of the study – no baseline data is given to show they were menstruating at base line?). The age given does not say if it is base or end line. If it is end line this means girls defined as 10-11y not have been 8-9y at baseline? Is this in the protocol – surely they were not menstruating at 8 y of age? Where has the control arm of the trial / study gone? Page 15 (Results; Table 1). Table should include 'are you menstruating', 'how long have you been menstruating' etc. in the text they state the average age of menarche as being 12.82y – this is extremely young – how was this question asked – it is not in the study measures and must be included in full, as this is a MHM study. Again age is confusing. Authors need to state age at end line if the results are end line. One assumes it is an error and authors are using baseline age and end line measures? Are the authors saying that 48 girls aged 10-13y at end line have all been menstruating throughout the follow-up trial / study? As age at menarche is asked at baseline for all girls, then it is simple to calculate how long they were menstruating in the study. Even girls of 10-13y at baseline line would very unlikely have been menstruating at baseline. This is a critical point as maybe some of these girls only menstruated at the end of the study time – how can their 'experience' be compatible with other girls aged say 19y who are truly menstruating. Again this reiterates the need to have baseline on all girls, perhaps as a

	supplementary table (see above, differences between transfers and trial participants). Table 1 is also confusing regarding going to same school in past year – the authors have stated all transfers were soon after baseline – but here at end line 11 girls have recorded they were at a different school. As with the question on all transfers, were they 'treated' exactly the same as baseline participants regarding consent, assent, baseline measures, provision of intervention, training how to use intervention? In survey measures, authors describe personal and MHM characteristics at follow up. It would be useful for a table of these same characteristics at baseline, to examine if there was change over time. This becomes particularly important since the study population seems to change over time, with transfers from others schools being included in the intervention arms. Page 16 (Results; Table 2): the breakdown by arm is needed since why would girls in the education arm be using AfriPads? Is there a reason control girls are not included in this study? Page 17 (Results; Table 3): This is intent to treat analysis that includes transfers. I am worried that the AfriPads would have been old and worn, or not given to the girls, yet they are all called AfriPad users. Why is this not presented as AfriPad, AfriPad plus education, education, control, as per the trial ? Surely AfriPad alone is different from AfriPad plus education? Table 2 and Table 3 seem to not correspond. 35.8% state they are using AfriPads, while Table 4 states 100% have access to clean absorbent. Maybe Table 2 should be separated into AfriPads v other? As the authors conclude the majority of girls do not use or like AfriPads, more information on the training techniques on use by the NGO is needed in the methods section. This becomes important as the implication for programmes is that reusable pads have no value above usual practice. It is thus essential to know if this is despite provision under the best possible conditions. Are the authors confident that girls were correctly taught? There is no mention of quality control or follow-up screening to check girls knew how to use the pads, and whether further training was given to ensure this. Reviewer 2 covered many questions specific to different MHM components. My concern is that poor quality interventions may mask any real differences between groups (and inadequate baseline prevents comparison over time). The discussion has been well edited. The sentence which recognizes the potential of insertable products should include studies that have actually tested the products in schoolgirls, notably Oster and Thornton 2011 in Nepalese schoolgirls, and Mason et al 2014 in Kenyan schoolgirls. Further issues raised in this review need to be considered in limitations.

REVIEWER	Bethany A. Caruso Emory University, USA
REVIEW RETURNED	22-Aug-2016

GENERAL COMMENTS	The authors have greatly strengthened this work by providing thoughtful and detailed clarifications of the text, particularly in the

	methods section. Before final acceptance, I do believe there is one question that warrants a second look as no changes were provided in the text. The question and response are pasted below: "D.3. p. 19, line 50. Attendance: In this paragraph, it is imperative that the researchers make it clear that this measure of attendance being assessed is general attendance and is not specific to attendance during a girls' menstrual period. Readers would have to be savvy to understand this important distinction as it is not made clear enough in the results or the methods. However, the distinction is critical, and the conclusions that can be drawn from any associations made are very very limited because this is not menstruation specific attendance data. I encourage the authors to be more clear and re-think the manner in which this association is discussed in this section. Response: We have attended to this in text and made this explicit in the methods. However, again would note that this has been standard practice in MHM studies and don't think that it would be surprising to readers that this was general attendance. " The authors are making an assumption of the readers of their work to interpret a vey critical measure. Not only do I not think readers will be savvy enough to understand the nuance, I believe that those who have read the other studies they are considering to be a part of 'standard practice in MHM studies' have misinterpreted the results shared. Moreover, just because a methodology for capturing attendance data has been the norm, does not mean that it is accurate (particularly if left open for interpretation). So to re-iterate, I think it imperative that the authors indicate that overall attendance data is not specific to when girls were menstruating in the LIMITATIONS section. The attendance data captured does provide some insight, however is limited. The specific association between MHM and attendance at school during menstruation is not measured and remains unknown.

VERSION 2 – AUTHOR RESPONSE

RESPONSE TO REVIEWERS

Reviewer(s)' Comments to Author: Reviewer: 1
Reviewer Name: PA Phillips-Howard
Institution and Country: Liverpool School of Tropical Medicine
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below
The authors have attended to most of the reviewer comments, however, there are a number of clarifications and edits now raised after the authors report this study population is part of the trial, and not an independent study.

Response:
Thank you for taking the time to provide additional comments on the manuscript.
While the schools and survey data were collected as part of the trial, this paper does not report on results of the trial. Menstrual hygiene was not an outcome in the trial nor have we compared it across

trial conditions. Rather, the secondary use of this final survey data to provide a prevalence estimate of MHM consistent with the definition, and to test associations between this definition and self-reported outcomes is a valuable stand-alone endeavour.
In revising the paper, we have noticed a few places where it was unclear that we compared girls USING AFRIpads provided in the trial (not comparing between the AFRIpad conditions). This was clear in the methods, but we have amended wording throughout other sections to be more consistent and to ensure this is clearer. We have also attended to this in abstract and in the conclusions of the paper to ensure the language is clear that the prevalence of MHM is only compared between those who self-report using different absorbents. (MHM is then used to predict outcomes, which is a substantial portion of the paper, and again is secondary-analysis of the survey data rather than any comparison of trial outcomes). We have added a section in the limitations to note that this study cannot speak to the effectiveness of the intervention only the difference between using/not using the AFRIpads with regard to the MHM prevalence – and that the associations with MHM aspects and hygiene are purely correlational.

Pg. 24 “The study compared girls who reported using AFRIpads provided in the trial as their primary absorbent to those using other methods when calculating MHM prevalence. While this presents a cross-sectional assessment of menstrual hygiene when using a reusable product, it does not reflect the effectiveness of providing reusable pads in an intervention. Girls’ were not compared across the intervention arms, rather across the primary absorbent used. Those in the reusable pad arms who may have received AFRIpads but not used them, were grouped according to the absorbent they used most often”

We believe this will address the reviewer’s concerns below, in that this study does not provide any comparison of trial conditions or outcomes of the trial (please see the research questions covered pg.7). Rather, compares girls cross-sectionally based purely on self-report. (The trial is relevant because that’s is how those using AFRIpads got them). As such, we do not imply that the finding that MHM did not differ between the two absorbent types meant that the AFRIpad conditions in the trial were ineffective, what we report is that for those who REPORT USING AFRIpads as their regular absorbent, their MHM does not differ. This is unsurprising given providing girls with a clean absorbent doesn’t address other aspects of MHM.
We have clarified this in multiple places throughout the manuscript.

Where the prevalence of MHM is compared – this is between girls who self-reported that AFRIpads were the absorbent they usually used, and girls who reported primarily using a different absorbent. This includes girls across all conditions (although only girls who received AFRIpads in the trial reported that they used this absorbent).
As such, we hope that the reviewer will understand that the trial protocol, baseline characteristics and other details, while all highly relevant for reports of the trial outcomes or comparison across trial arms, should not be required for this paper. Further, we feel they would be counterproductive to the goals of the reviewer and authors – we want to ensure that readers do NOT confuse this paper with the results of the trial and that they understand that this is a cross-sectional study (with all the inherent bias of that approach!).

Main comments
There is no publication for the trial reference (Ref 24), which the authors note is taking time, but it is frequently used to cover elements of this study. The authors state this study is not separate from the trial, and data are predicated on the trial methods with the design aiming to assess relationships with outcomes compared across arms (and the title reports ‘impact’). It is thus important that more detail on the trial methods is provided, it remains insufficient for a stand-alone paper comparing outcomes. It does not matter that there is duplication.

Response: We have undertaken additional revision of the paper to ensure clarity around the trial methods and the relationship of the trial to this study.

We are also frustrated by the delay in publication of the trial paper, but it is not the case that we have used this to cover up elements of this study.
As noted above the data was collected as part of the trial but this paper does not report trial outcomes. The statement that the title claims the study reports 'impact' is misleading. The title implies a measurement of the impact of poor menstrual hygiene (in a cross-sectional comparison with other outcomes) – not of an intervention, and not of AFRIpad provision. The paper reports cross-sectional analyses that compare girls' self-reported practices against their health and education outcomes. More focus on the trial would take away from the primary aims of the paper which are (1) to measure menstrual hygiene in a way consistent with the concept definition, and (2) to see if menstrual hygiene (as defined) is associated with health, education and wellbeing – in this survey sample.

Examples of methods requiring more detail include the sample size calculations, inclusion/exclusion criteria for recruitment of the girls and recruitment of the schools, baseline measures of participants, and measures on intervention rigor.
A copy of the trial protocol is also required to accompany the review of this submission, and as an online supplement, as results are comparative between arms.
I recommend removing the late entrant (transfers). If the authors are to include the 60 transferred girls, not enrolled at baseline they need to better characterize the girls to show they are the same – it looks as if they have added them just to boost numbers, without indicating if these girls were 'treated' exactly the same regarding interaction with research and implementing staff, training, provision of interventions etc. This is discussed further below, particularly regarding intervention allocation when conducting the trial / study ITT analysis.
Data on actual menstruation (both base line and end line measures) needs clarification, especially with reported inclusion of very young girls (10-11y olds) at endline. Menstruation must have been an eligibility criteria to enter the trial / study. Girls who were not menstruating at baseline should be excluded or only included once they start menstruating.

Response:
We have added a reference to the trial registration.
We stress again that this paper does not provide an analysis of the trial outcomes. There is no intention to treat analysis presented here. Moreover, girls are never compared according to condition – they are compared cross-sectionally according to what absorbent they are actually using based on the survey alone.
As noted above, we have revised language throughout to ensure this is clear and to remove anything that could be interpreted as suggesting that the trial was responsible for differences beyond the different absorbent used.

We have sought to again emphasise in revisions that of the transferred girls who received the intervention this was exactly the same as those 'officially' in the trial (ie. Tracked from baseline).

Menstruation was not an eligibility criteria for the trial. We have clarified in revisions. Menstruation is an eligibility criteria for inclusion in this secondary analysis as questions about absorbent etc were only asked of those who were menstruating.

The strengths and weaknesses section on the title page needs amending to take into account the weaknesses in this trial / study, noted below. The current first and second (and perhaps also third) can be combined, providing room for reporting the weaknesses. The main weakness are detailed below, including inclusion of non-trial girls due to small trial sample size and intervention allocation, poor school attendance measures, information collected on menarche and menstruation (see

comments on result, Table 1).

See responses below. This was not a trial study.

Other comments
Page 2: the conclusions on AfriPad may not be accurate – more detail is needed on whether girls received the AfriPads, whether they were trained adequately, whether the additional 60 girls received intervention exactly the same as the 145.

Response: Again we aren't reporting the effectiveness of the intervention – just the use of AFRIpads – this would be similar to if girls had purchased the AFRIpads themselves. It is not a reflection on the results of the trial. We have revised to ensure this is clear.

All girls provided with the AFRIpads received the exact same intervention. As per our last response to reviewers, many of these additional girls were in the schools from the end of the baseline attendance data collection period – meaning they received the trial exactly the same as the included girls -but didn't have baseline attendance data.

Page 7 (Survey design): The design is an end line cross-sectional survey nested within a trial. why are the controls missing from this trial / study? Were late entrants (transfers) included in the design?

Response: The controls are not missing from this study. They are included alongside the girls not using AFRIpads. We would urge the reviewer to consider the purpose and story of this paper, not the trial, which is not reported here.

Page 7 (Sample size calculation): STROBE wrongly checks this as presented. With due respect to the authors response, sample size calculations are needed. They can be the sample size calculations in the trial protocol, which aim to measure the outcomes reported in this paper. It is clear this is a limitation (but if correctly presented does not preclude publication) and should be included as a weakness of the study in the strengths and weaknesses section in the title page.
Page 7 (Intervention methods): The new edits bring up some questions on trial interventions which are relevant to participant responses to MHM. A section in methods on how the interventions are provided is required. Attention to how it was provided to the 29% of sample who are transfers is also needed (including the 11 in Tale 1 who were at a different school the previous year).

Response: Again we note that this paper uses the follow-up survey data but does not report trial results.
As noted on the title page of the paper, the study and analyses are limited by the small sample size, and lack of existing literature on which to base power analyses. We do not believe that the sample size calculation from the trial is relevant to this paper as it was powered based on past findings for attendance. That is not the primary outcome in this study, nor is there sufficient background literature to base a power analysis on for this study where we compared self-reported menstrual hygiene with outcomes (e.g., we compare washing/drying practices and outcomes).

While the authors state the AfriPads are not the main outcome, a major finding in the abstract and results relate to comparison between arms, showing no difference in MHM between girls in the AfriPads and control arms. It thus remains important to clarify whether AfriPad provision was adequate for these analyses, both in terms of optimal provision, training, and the proportion of girls within this arm receiving the pads. One AfriPad pack containing 2 reusable pads was given in October 2012, and then nothing until March 2014 – this means that girls had 2 reusable pads to last for 18 months. This seems very meagre, other reusable pads are stated to last 4-6 months before disintegrating. Can the authors confirm that AfriPads have a 'shelf-life' of 18 months and remain fully

effective for MHM? If not this needs to be discussed in the limitations and discussion, as it would seem obvious that girls will not continue to use old, poor quality materials, resulting in poor MHM (i.e. an inadequate intervention adds to poor MHM, rather than resolving it). This should also be added as a weakness in the strengths and weaknesses section in the title page.

Response: This paper does not compare different arms of the study. It calculates the prevalence of menstrual hygiene separately for those who report using AFRIpads as their primary absorbent as they received this in the trial. We have revised the abstract and conclusions of the paper to ensure this is clear and to give less prominence to this in those sections, as it was not the primary or secondary aim of the study to draw this comparison.

We should also note to the reviewer that AFRIpad delivery included a PACK of AFRIpads – as noted in the information provided about the trial (pg. 6) this includes 'base' liners (which affix to underwear), onto which are placed 3 straight and 3 winged liners. Thus, functionally a total of 6 pads were provided at each delivery. AFRIpads can be used for at least 12 months (www.afripads.com). We have added this note to the text.
We would note that this study compares girls currently using the AFRIpads to those using other methods – the AFRIpads they would be using were provided in March 2014 – as noted.

please confirm menstruation is an eligibility criteria.

Response: Menstruation is not an eligibility criteria for participation in the trial, we have included this information in revisions. Menstruation is a criterion for inclusion in the secondary analysis of the survey data presented here (only menstruating girls were asked relevant follow-up questions in the survey). As noted in the trial section of the introduction and in the methods – all girls were surveyed to avoid potential harms of identifying only menstruating girls for survey. Something we feel is very important to the ethical conduct of trials in this field.

The new edits state that girls were provided the AfriPads in the intervention group on recruitment, and received training to use them. Was this conducted at one time-point for all girls? It is much less clear if girls transferring into the school received the intervention with the same training activities. The authors note that the trial analysis of outcomes between arms is intention to treat analysis – which is usually based on provision at enrolment after screening eligibility. If new entrants come later to a cluster, are these girls classed in the trial as eligible for enrolment through rolling enrolment? It is assumed they are included to increase sample size. If they arrive (transferred) from another study school they usually hold allocation for their previous school. This trial appears to allocate the new entrants in the intervention of the new school. The authors need to confirm they are similar in terms of allocation and training; otherwise they would dilute the prevalence of use at the final survey, and reduce the impact of the intervention. The number who transferred after the baseline is not a small portion, with 60 of the 205 girls recorded as late entrants (29%). While the limitations section says 'many' of these were given the intervention; the number and portion is not given – authors need to specify how many received the AfriPads and the correct training. Further, how do authors know the late entrants had the same characteristics at baseline – what measures do they have to confirm this?

Response: Intention-to-treat analysis means all girls who were recruited into the study at baseline. This includes girls who did, and did not, receive the intervention. There is no intention-to-treat analysis in this paper. We have revised the mention of this in the methods in case it is confusing.
Again, we must stress that analyses in this paper are cross-sectional, we compare girls who report using AFRIpads, with girls who report using other methods. This is irrespective of condition (although only those who received AFRIpads in the intervention were using them, no girls reported having purchased them etc.). Most analyses compare other self-reported behaviours.

Similarly, it is unclear whether any of the 60 girls recruited later who entered education arms received the same puberty education that the NGO provided as standard to all girls after baseline observations in all girls. Can the authors verify this? If not, the value of including these additional girls is doubtful. Unless there is strong evidence of similar allocation for late entrants, it would be better for them to be excluded or be classed as controls to prevent them diluting the intervention effect. While more detail on this is required, these issues need to be discussed in the 'limitations' section, and possibly as a weakness in the strengths and weaknesses section in the title page.

Response: As reported in the methods – 2 of the transfer girls received the education. This means that they were in the exact same education sessions as the trial girls.

This paper does not report on any comparisons between the trial conditions. The only comparison for which the trial is essential is between girls currently using the AFRIpads, and girls who are not using the AFRIpads, this is not segregated according to the intervention – but according to girls' self-reported absorbent use. Therefore, transfer girls who did not receive the AFRIpads are grouped with 'controls', and transfer girls who received the AFRIpads alongside the trial girls are included in this group.
We have added a note about this in discussion but hope that the reviewer will find this clear in revisions.

Was the second provision at 18 months at the finish of the trial? Was this to controls or just the pad users – was there an improvement in use once they received the new pads?
Page 9 (Survey measures): Measures of baseline characteristics required.
Page 10 (Survey measures; health): there are questions on dealing with periods but authors need to include the (biological) menstrual questions i.e. have you ever menstruated (Y/N), when did your menstruation start (date), is your menstruation light or heavy etc. this needs to be reported at baseline and end line. My worry is, that girls not menstruating may slip through and answer 'No' for questions on dealing with their period (i.e. did you use AfriPads – No).

Response: The second provision was for schools in the reusable pad provisions. We feel this is clear in the section describing the interventions.
See note at the start of our response.
The above questions mentioned were included in the present study (although not in raw form) – girls who had menstruated were included in the study (trained research assistants noted any cases where they may have suspected girls answered a question wrong, or questions were inconsistent – considerable cleaning was undertaken on the data set to ensure consistency).
The start date of menstruation was used to calculate age at menarche (deducted from the date of survey).

Authors state all the 205 girls reached menarche – was this at baseline? Can the authors confirm if the eligibility criteria was all menstruating girls at baseline? Is this in the protocol? Ages are confusing for endline measures – see below on results comments. If girls who were not menstruating at baseline are included this needs to be clearly stated. Surely 10-13 year olds at endline would not be menstruating at trial / study start, and likely a high proportion of these would only just have begun menstruation toward the end of the study; such girls would have little experience of MHM practicalities, so would be different and dilute the outcomes. At the moment only follow-up data is mentioned; age at menarche at the beginning of the study would be useful as it would allow exploration of menstrual experience which many explain differences in ability to deal with MHM and also use of AfriPads. (It appears this was collected from trial girls at baseline - as age at menarche is noted at the beginning of results – was it collected also from the additional 60, so all 205 girls?).
Maybe girls who had been menstruating for a number of years were better able to deal with AfriPads compared with girls who started menstruating just before the trial began?

Other characteristics measured for end line are very thoroughly reported so it is strange to miss the menstrual questions which are at the heart of the trial study.

Response: No baseline data is reported here. This study is very much a stand-alone paper that is a secondary analysis of the survey data collected at follow-up.
We do not compare intervention conditions, we do not compare longitudinally over the study.

As the reviewer notes, we thoroughly report the questions that are relevant to these analyses, and do not confuse this study with the questions specific to the trial outcomes.

Page 11 (Survey measures; school attendance): elsewhere it states school registry was used for assessing absence at baseline for the trial. Here the authors switch and say for the present study (but the authors have said it is not a separate study, but same as the trial). Baseline attendance using registers would not have captured the 60 transferred girls added to the 145 trial participants. The authors say they have responded to Reviewer 1 and 2 regarding the attendance spot-check – stating that it is common to use general attendance, not menstrual specific. They have not dealt with the more important concern of switching methods of measuring attendance, or the ability of spot-checks rather than continuous surveillance to adequately document this important outcome. In the discussion they state it is not logistically feasible to do continuous attendance monitoring. Use and reliability of spot-check needs to be reported as a weakness in the strengths and weaknesses section after the title page. Non-collection of absence in menstruation is also a weakness.

Response: We apologise if the past response to reviewers added confusion. The survey was not conducted as a separate study – but this paper is.

We state earlier in the paper that “attendance registers” were taken at education sessions (ie. The puberty education). Nowhere in this paper is it stated how attendance was collected at the baseline – so we assume the reviewer has got this information elsewhere. Changes in attendance data are very relevant for the primary trial outcomes and are discussed at length where they are relevant. Baseline data is not used at any point in this paper (other than in the introduction to say schools were comparable at baseline), so there is no comparison to be made between baseline and follow-up attendance data.
Again we are concerned that the reviewer may have confused this paper as reporting on the outcomes of the trial, which is not the case. The only way that attendance data is used in this paper is that endline attendance (all collected consistently) is cross-sectionally compared with self-reported menstrual hygiene behaviours.
At the request of reviewers we have further highlighted that this paper reports includes a measure of general attendance, not menstrual-specific attendance, but disagree with both reviewers that this is a pronounced weakness. Rather, we feel there would be both pros and cons of menstrual-specific attendance given these would rely on self-reported menstrual calendars. We have expanded this section of the discussion significantly – see response to Reviewer 2.

Page 12 (Analysis): In the authors’ response to R2, they state SDQ analysis was conducted in detail in the trial paper – is this the same data? The full description is needed as this is a standalone paper. Is all analysis like the trial, using intent to treat – is that in the protocol?
Inclusion of transferred girls pooled with the trial girls seems still to be only done to increase sample size. Have the authors conducted analysis without the transferred girls? a supplementary table is needed to show these girls ‘are the same’, which the authors state. This should include a comparison of baseline and of end line characteristics.

Response: Again we note that this study provides no comparison between trial conditions, and no analysis of trial outcomes, only cross-sectional analyses.

We have included more detail about the SDQ in methods, including typical cut-points etc.
We are struggling to understand how this paper could have been interpreted as an intention to treat analysis.
In this study we:
- Calculate a prevalence estimate for menstrual hygiene based on the survey information provided
- We calculate this separately for those who were using AFRIpads as these were provided in the trial and would not have naturally be in the population. We note that this prevalence does not differ – in revisions we have tried to be as clear as possible that this was not compared between trial conditions and compares those USING AFRIpads
- We then use the aspects of menstrual hygiene to look at the associations between aspects of menstrual hygiene and outcomes. As predictors we use: clean absorbent (Where we consider AFRIpads alongside clean cloth & sanitary pads), washing practices, drying practices, and our estimate of privacy. These are all cross-sectionally compared in their associations with health, education and psychosocial wellbeing. This collapses across all conditions in the trial and compares girls based on their self-reported MHM behaviours.

Page 13 (Ethics): The authors are requested to provide the protocols. Was the process of consent from the parent and assent from each girl the same procedure for transferred girls – it states girls at the start of the study gave verbal consent – is that different from the written consent they had already given? Again would this miss transferred girls?
Pages 14/5 (Results; participant characteristics, Table 1): this appears to be end line participant characteristics. The first sentence is two sentences – the first half states 205 girls were menstruating (is this at the beginning or at the end of the study – no baseline data is given to show they were menstruating at base line?). The age given does not say if it is base or end line. If it is end line this means girls defined as 10-11y not have been 8-9y at baseline? Is this in the protocol – surely they were not menstruating at 8 y of age?
Where has the control arm of the trial / study gone?
Page 15 (Results; Table 1). Table should include ‘are you menstruating’, ‘how long have you been menstruating’ etc. in the text they state the average age of menarche as being 12.82y – this is extremely young – how was this question asked – it is not in the study measures and must be included in full, as this is a MHM study.

Response: Please see above (first response). No baseline data is included. To include any baseline data in this study would greatly confuse readers, and – as above – it is really not relevant to the purpose of this study.
Age at menarche is calculated from the date of menarche provided in the questions (as noted by the reviewer above). The self-report nature of the data is a limitation, we have noted this in the discussion.

Again age is confusing. Authors need to state age at end line if the results are end line. One assumes it is an error and authors are using baseline age and end line measures? Are the authors saying that 48 girls aged 10-13y at end line have all been menstruating throughout the follow-up trial / study? As age at menarche is asked at baseline for all girls, then it is simple to calculate how long they were menstruating in the study. Even girls of 10-13y at baseline line would very unlikely have been menstruating at baseline. This is a critical point as maybe some of these girls only menstruated at the end of the study time – how can their ‘experience’ be compatible with other girls aged say 19y who are truly menstruating. Again this reiterates the need to have baseline on all girls, perhaps as a supplementary table (see above, differences between transfers and trial participants).
Table 1 is also confusing regarding going to same school in past year – the authors have stated all transfers were soon after baseline – but here at end line 11 girls have recorded they were at a different school. As with the question on all transfers, were they ‘treated’ exactly the same as baseline participants regarding consent, assent, baseline measures, provision of intervention, training how to

use intervention?
In survey measures, authors describe personal and MHM characteristics at follow up. It would be useful for a table of these same characteristics at baseline, to examine if there was change over time. This becomes particularly important since the study population seems to change over time, with transfers from others schools being included in the intervention arms.

Response: No baseline data is reported in this paper as it is a secondary analysis of the follow-up survey. Age is self-reported. In this sample some girls were unsure of their age, and while we do believe that the reported age reflects genuine self-report, we also don't expect this to be 100% accurate. Again, since we compare girls according to their menstrual practices we believe all experiences are relevant. This is a cross-sectional assessment, we compare girls on what they are CURRENTLY doing, and their CURRENT symptoms/outcomes.

The population does not change over time for this study. Girls were all surveyed at the end of the trial, some of the girls who did not have baseline data were also surveyed and included here (thus only lack the attendance data for that one comparison), some of these girls had also received AFRIpads as part of the trial and so reported using AFRIpads as their primary absorbent at follow-up. Girls are then compared in their level of menstrual hygiene between those who report using AFRIpads and those who report using another method.

Page 16 (Results; Table 2): the breakdown by arm is needed since why would girls in the education arm be using AfriPads? Is there a reason control girls are not included in this study?

Response: This study is a secondary analysis of follow-up data. Girls are compared cross-sectionally according to their menstrual hygiene practices, not according to conditions. Girls from all conditions are included in the study.

Page 17 (Results; Table 3): This is intent to treat analysis that includes transfers. I am worried that the AfriPads would have been old and worn, or not given to the girls, yet they are all called AfriPad users. Why is this not presented as AfriPad, AfriPad plus education, education, control, as per the trial ? Surely AfriPad alone is different from AfriPad plus education?
Table 2 and Table 3 seem to not correspond. 35.8% state they are using AfriPads, while Table 4 states 100% have access to clean absorbent. Maybe Table 2 should be separated into AfriPads v other? As the authors conclude the majority of girls do not use or like AfriPads, more information on the training techniques on use by the NGO is needed in the methods section. This becomes important as the implication for programmes is that reusable pads have no value above usual practice. It is thus essential to know if this is despite provision under the best possible conditions. Are the authors confident that girls were correctly taught? There is no mention of quality control or follow-up screening to check girls knew how to use the pads, and whether further training was given to ensure this.

Response: This is not an intention to treat analysis. This is a cross-sectional analysis of survey data. The only girls considered in the AFRIpad group are those who report using AFRIpads as their primary menstrual absorbent. So the AFRIpad group includes ONLY those actually using the absorbent. It is not a comparison according to conditions.
Table 4 does not report on absorbents at all? Only on health, education and psychosocial wellbeing outcomes?
The authors never conclude that girls "Do not use or like AFRIpads". This paper never compares across conditions. This paper is not a comparison of the effectiveness or impact of AFRIpads – this paper compares self-reported menstrual hygiene practices to outcomes. With the exception of calculating menstrual hygiene (as per the concept definition) where AFRIpads are noted separately because they were provided in the trial, this study doesn't compare AFRIpad use on other outcomes. The reviewer is directed to the main analyses presented in Table 5 – which groups AFRIpad with

clean cloth for predicting health, education and psychosocial wellbeing outcomes.

Reviewer 2 covered many questions specific to different MHM components. My concern is that poor quality interventions may mask any real differences between groups (and inadequate baseline prevents comparison over time).

Response: We direct the reviewer to Table 5 which reports on the final study analysis. In these analysis we compare girls' self-reported behaviours with their self-reported consequences for menstrual hygiene.
The contribution this paper makes is in measuring menstrual hygiene and comparing different aspects to self-reported outcomes. In this way the first review provided was incredibly useful to improving the paper – but we note here that no comparison is provided across-group and we aren't evaluating the effectiveness of the intervention in this paper at all.
We have revised to ensure we have discussed the training provided with the AFRIpads when discussing this result, but it is not the main focus of the study.

The discussion has been well edited. The sentence which recognizes the potential of insertable products should include studies that have actually tested the products in schoolgirls, notably Oster and Thornton 2011 in Nepalese schoolgirls, and Mason et al 2014 in Kenyan schoolgirls. Further issues raised in this review need to be considered in limitations.

We have provided some further edits to the discussion as noted above to ensure clarity around the purpose and methods of this study, we have included the references to menstrual cup studies.

Reviewer: 2
Reviewer Name: Bethany A. Caruso
Institution and Country: Emory University, USA
Please state any competing interests or state 'None declared': None Declared
Please leave your comments for the authors below The authors have greatly strengthened this work by providing thoughtful and detailed clarifications of the text, particularly in the methods section. Before final acceptance, I do believe there is one question that warrants a second look as no changes were provided in the text.
Response: Thank you for taking the time to re-review this paper. We agree that the past comments greatly strengthened the work and are grateful.

The question and response are pasted below:
"D.3. p. 19, line 50. Attendance: In this paragraph, it is imperative that the researchers make it clear that this measure of attendance being assessed is general attendance and is not specific to attendance during a girls' menstrual period. Readers would have to be savvy to understand this important distinction as it is not made clear enough in the results or the methods. However, the distinction is critical, and the conclusions that can be drawn from any associations made are very very limited because this is not menstruation specific attendance data. I encourage the authors to be more clear and re-think the manner in which this association is discussed in this section.
Response: We have attended to this in text and made this explicit in the methods. However, again would note that this has been standard practice in MHM studies and don't think that it would be surprising to readers that this was general attendance. "
The authors are making an assumption of the readers of their work to interpret a very critical measure. Not only do I not think readers will be savvy enough to understand the nuance, I believe that those who have read the other studies they are considering to be a part of 'standard practice in MHM studies' have misinterpreted the results shared. Moreover, just because a methodology for capturing attendance data has been the norm, does not mean that it is accurate (particularly if left

open for interpretation).
So to re-iterate, I think it imperative that the authors indicate that overall attendance data is not specific to when girls were menstruating in the LIMITATIONS section. The attendance data captured does provide some insight, however is limited. The specific association between MHM and attendance at school during menstruation is not measured and remains unknown.

Response: We are curious that others have misunderstood the attendance measures in past papers – that is indeed unfortunate. We certainly agree that standard practice does not necessarily mean the method does not have limitations. We added discussion of this in the limitations section as requested.

While it is true that the direct effect between MHM and attendance remains unknown we think there would be additional issues in using a 'menstruation-specific-attendance' outcome. This would introduce bias at girls' self-reported cycle and likely result in a lot of missing data where girls lose or do not record menstrual cycles (as well making quite a significant request of participants to record their cycle and deliver to the researchers). In contrast, general attendance data does dilute the impact of the predictor (for example in this instance we are comparing according to say, absorbent type, where a menstruation-specific measure would have a maximum of 100% menstrual days with one type, and 0% menstrual days with another, our data dilutes the effect size because even if girls regained 3 or 4 days due to a different absorbent this would show up as ~11% difference). It also means that confounds that may influence general attendance and the type of absorbent used may have a greater role. We have sought to note this balanced appraisal of the limitations (and also strengths) of using general attendance (objectively collected, rather than self-reported) in the limitations section.

VERSION 3 – REVIEW

REVIEWER	Penelope Phillips-Howard Liverpool School of Tropical Medicine, UK
REVIEW RETURNED	30-Oct-2016

GENERAL COMMENTS	The authors have clarified most of the issues and concerns raised in the first review, with appropriate amendments made to the paper. A few minor comments remain, as below. The authors report that there was only a time difference of 7-8 months between last provision of AFRIpads and the survey, and that AFRIpads should last for 12 months. It would be valuable to add this information into the first or the last paragraph of the discussion as it suggests deterioration of the pads were unlikely to be a contributory factor. It is unclear why reference 26 is only used when there are research papers on insertable menstrual items such as cups among African women and girls, as well as the Oster and Thornton paper examining the cup among schoolgirls in Nepal, including: -Beksinska ME, Smit J, Greener R, et al. Acceptability and performance of the menstrual cup in South Africa: a randomized crossover trial comparing the menstrual cup to tampons or sanitary pads. J Womens Health (Larchmt) 2015; 24(2): 151-8. -APHRC. Policy Brief: Attitudes towards, and acceptibility of, menstrual cups as a method for managing menstruation. Nairobi: African Population and Health Research Center; 2010. -Mason L, Laserson K, Oruko K, et al. Adolescent schoolgirls' experiences of menstrual cups and pads in rural western Kenya: a

	qualitative study. Waterlines 2015; 34(1): 15-30. -Tellier M, Hyttel M, Glad M. Assessing acceptability and hygienic safety of menstrual cups as menstrual management methods for vulnerable young women in Uganda Red Cross Society's Life Planning Skills Project. Kampala, Uganda, 2012. -Averbach S, Sahin-Hodoglugil N, Musara P, Chipato T, van der Straten A. Duet for menstrual protection: a feasibility study in Zimbabwe. Contraception 2009; 79(6): 463-8.

VERSION 3 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Penelope Phillips-Howard
Institution and Country: Liverpool School of Tropical Medicine, UK
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
The authors have clarified most of the issues and concerns raised in the first review, with appropriate amendments made to the paper. A few minor comments remain, as below.

RESPONSE: Thank you for taking the time to re-review the paper, and for your useful comments throughout the process which have strengthened the manuscript.

The authors report that there was only a time difference of 7-8 months between last provision of AFRIpads and the survey, and that AFRIpads should last for 12 months. It would be valuable to add this information into the first or the last paragraph of the discussion as it suggests deterioration of the pads were unlikely to be a contributory factor.

We have added this comment in the discussion - page 22. We also note there that AFRIpad care may have varied and we don't know how well girls maintained the pads.

It is unclear why reference 26 is only used when there are research papers on insertable menstrual items such as cups among African women and girls, as well as the Oster and Thornton paper examining the cup among schoolgirls in Nepal, including:

RESPONSE: Thank you for the below suggestions. We have added the additional citations [new refs: 32-25].

-Beksinska ME, Smit J, Greener R, et al. Acceptability and performance of the menstrual cup in South Africa: a randomized crossover trial comparing the menstrual cup to tampons or sanitary pads. J Womens Health (Larchmt) 2015; 24(2): 151-8.
-APHRC. Policy Brief: Attitudes towards, and acceptibility of, menstrual cups as a method for managing menstruation. Nairobi: African Population and Health Research Center; 2010.
-Mason L, Laserson K, Oruko K, et al. Adolescent schoolgirls' experiences of menstrual cups and pads in rural western Kenya: a qualitative study. Waterlines 2015; 34(1): 15-30.
-Tellier M, Hyttel M, Glad M. Assessing acceptability and hygienic safety of menstrual cups as menstrual management methods for vulnerable young women in Uganda Red Cross Society's Life Planning Skills Project. Kampala, Uganda, 2012.
-Averbach S, Sahin-Hodoglugil N, Musara P, Chipato T, van der Straten A. Duet for menstrual protection: a feasibility study in Zimbabwe. Contraception 2009; 79(6): 463-8.