Reviewers report

Title: Vaginal douching in Zambia: A risk or benefit to women in the fight against cervical cancer: A retrospective cohort study

Version: 1 Date: 09 May 2019

Reviewer: Katie O'Brien

Reviewers report:

Vaginal Douching and Cervical Cancer in Zambia

I read this article with great interest, as this is an understudied topic and it seems particularly relevant to a country like Zambia where both douching prevalence and cervical cancer incidence are very high. Overall I liked the article and I think the findings are very interesting and worthy of publication. I thought the authors did a great job of fitting their work into the greater literature and have a fair interpretation of what their results mean. However, I do have some concerns that I would like to see addressed.

Major concerns

1. I would have liked to see more information on the study sample. I know that it has been published on before, but the manuscript is missing a lot of key information, including how information on douching exposure and the covariates of interest were assessed, including the timing relative to when the VIA was performed. I also would have liked to have more information on who attends the screening clinics so that I could better understand how this sample is representative of the Zambian population. I additionally noticed that one key eligibility criteria - having at least one sexual partner - was not adequately described in the methods, which leads me to question what else may have been omitted.

2. The authors examine what type of douching is associated with cervical lesions, but never give us estimates for ever versus never, which I think is of primary importance. I also expected to see never douching as the referent group for the comparison of types of douche used. Point of clarification: douching with plain water is still douching if it involves a pressure system/ tubing, yes?
3. There are some concerning errors in the only full paragraph on Page 6. First of all, I don't think women older than 49 are included, yet age groups 45-54 and 55 or older are both mentioned. Secondly, the current interpretation of the education category %s is incorrect: It should be that among those who douched, 41% had a secondary education and 36% had a primary education. (Rather than the % who douched among those with a secondary education, as described.)

4. Tables 1-3 could easily be combined. Table 2 could even be eliminated entirely, as the numbers are directly calculable from the n's provided in Table 1 and the estimates are not even discussed in the results. Further, since this is an observational study, I would assume that the effect estimates are subject to some confounding and am unsure of how to interpret the crude estimates anyway. Finally, it is not necessary to include both the 95% confidence intervals and the p-values: the 95% CI contains all of the information in the p-value, along with extra information about the precision of the estimate.

5. For variables that are continuous (e.g. age, number of sex partners, age @ debut, income), ORs for the continuous form of the variable and/or trend tests would help with interpretation.

6. Table 5 could also be greatly reduced. Although the initial analysis is more about identifying what factors predict whether or not a woman douches, the second analysis is more directly about the association between douche and cervical lesions. As discussed above, I think the primary focus should be on ever vs. never douche with a secondary analysis on type of douching. The effect estimates for the other covariates are less interpretable in the current form as they are likely not appropriately adjusted for. For example, the association between age and cervical lesions cannot be confounded by any of the other factors in the table because nothing affects age. There may be other key covariates/confounders of the HIV-cervical lesion or occupation-cervical lesion associations that are not included here, making those difficult to interpret as well.

Minor comments:

• I noticed that a lot of women are missing HIV status. Normally I would discourage having a "missing" category in the tables/ as a covariate, as this results in a heterogenous category of people that it is not appropriate to perform statistical tests on as a group. However, since so many women are missing this information, I understand why excluding this women is not a reasonable approach. That being said, I think you need to acknowledge this as a limitation in the discussion. If you can think of a reasonable alternative (imputation, exclusion of this variable from multivariable analyses), that might be informative, please consider adding that as a sensitivity analysis.

• In the first paragraph of the results section, please provide %s as well as the n's. The %s are easier to make sense of.
An OR of 1.6 should be interpreted as 1.6 AS likely not 1.6 times MORE likely (see page 8).

Table 4 could be omitted. Model selection methods are inherently biased and it adds no extra information to what is already seen in the previous tables. If kept, it might be helpful to re-frame as an attempt to build a predictive model for douching, rather than trying to interpret each covariate estimate on its own.

The conclusions are very circumspect (e.g. "include messages to sensitize women against douching"). I'd prefer to see something a bit more direct (e.g. "public health messaging should describe the possible health risks of douching") while still avoiding causal language.

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable
**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal