Reviewer's report

Title: Prevalence and Associated Factors of Cervical Intraepithelial Neoplasia: A Cervical Cancer Screening Program Performed on 728,704 Women in Beijing

Version: 4 Date: 26 March 2014

Reviewer: OLIVER Chukwujekwu EZECHI

Reviewer's report:

Reviewer's report
Title: Prevalence and Associated Factors of Cervical Intraepithelial Neoplasia: A Cervical Cancer Screening Program Performed on 728,704 Women in Beijing
Version: 1 Date: 26th March 2014

Reviewer's comments

General comment: I appreciate the Editor for the opportunity to review this manuscript on a very important public health issue affecting women globally. I had great difficulty reading the manuscript as the quality of English was low, making the manuscript unfriendly to read. The manuscript requires extensive editing.

My specific comments to the authors are as follows:

Title page: The title of the manuscript is not concise and requires editing to clearly capture in one sentence what the study was about. Also the use of cervical intraepithelial neoplasia in the title is inappropriate as it is a histological diagnosis. However if the diagnosis is histologic following cervical biopsy it should be stated in the method section.

Two authors were designated as corresponding authors. This is not acceptable as only one author can take responsibility of the corresponding author task. However if the contribution of both authors are equally, it should be stated but only one should be designated the corresponding author.

Abstract: Need to be re-written especially the conclusion which should only capture in one or two sentences the main study finding and end with recommendation based on the finding.

Background: Is scanty and did not identify the gaps in knowledge or contribution to knowledge the present study is making. In the present state, it is disjointed and not organized. Paragraphs were not linked to each other. No mention of the aim of the study.

Methods section: The study design, subjects and setting were not clearly stated. The authors should clearly state them. The authors described that their study population were from different municipalities and need to describe the setting of each municipality clearly. In the result section they should compare the characteristics to show they are dealing with one population as difference may
impact on the outcome of their multivariate analysis.

Data collection: How the data were collected was not clearly stated as I was confused with the statement'………medical doctors conducted the examination blinded to the result of the TCT test". Who obtained the specimen? What was the logic in “blinding” the test after collecting the specimen? What was actually blinded?

The information the authors collected, what informed the choice- literature or experience. All collected exposure variables need to be defined and how were they defined/diagnosed need to be stated in a sub section to be titled definition of variables.

Outcome variable. I got confused here. How was the diagnosis made? Histologic or cytological. If histologic? , were cervical biopsy samples taken. If this was done is then not screening but diagnostic. I suggest the authors use the standard Bethesda nomenclature.

Statistical analysis: Though this is a manuscript with potential, however the analysis need to taken a bit further in order to fully take advantage of this potential. I think that this study has a unique opportunity to look at the validity of the observation SIL/HSIL common among some category of Chinese women . The question is whether that is a genuine fact, or if it could be explained by confounding variables. This data set has the probably unique opportunity to test this, since a number of such potential risk factor actually was recorded among the participants . Therefore I suggest the following:

1. Make a table where you test the bivariate association between each of the mentioned risk factors and the outcome variable SIL/HSIL expressed by Odds Ratios. This way it is possible to see if the first criteria for being a confounder is full-filled: Whether it is a predictor of the outcome.

2. In the next analysis, perform a step-wise logistic regression analysis, starting with the crude association and successively introduce those risk factors which were shown to be statistically significantly associated with the outcome in earlier analysis.

3. If the associations still would be statistically significant after the introduction of all potential risk factors identified earlier, then there is a very strong case that there is a biological cause for the associations. This additional step would be a unique contribution to the knowledge in this area, although I do not know for sure whether it has been done in any of the previous studies from china, but usually precisely this type of information on important confounders are most commonly lacking in studies which are based on register information or information derived from standard patient records. Pursuing the suggestions above would really make use of the full potential of the data set underlying this study!

Sample size determination: I have some concerns with sample size. How it was calculated was not stated.

Results: Need to be rewritten after the suggested analysis. The subject’s characteristics should come first before prevalence.

Discussion. Need to be rewritten, first stating the main finding in the first
paragraph and thereafter discussing each of the findings vis a vis existing literature.

Conclusion: Need to be written as it did not capture the study findings.

Authors’ contribution: Some of the contribution of the authors did not merit authorship and the list need to be revisited.

Reference: Almost all the reference did not follow BMC style and need to be rewritten to conform to the style.

Authors should check all the Tables. They are very difficult to understand.

Authors should review for clarity. Occupation should be classified using standard classification. Table 2, 3 and 4 should be merged into one

Recommendation: Major compulsory revisions

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: Yes, and I have assessed the statistics in my report.

Level of interest: An article of importance in its field

Quality of written English: Not suitable for publication unless extensively edited

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests'